

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:03:55 ; Search time 13.4328 Seconds  
(without alignments)  
198.395 Million cell updates/sec

Title: US-10-036-371-2

Perfect score: 107

Sequence: 1 IVGGYTCGANTVPYQVSLNS 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002:\*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*

2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*

3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*

4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*

5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*

6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*

7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*

8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*

9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*

10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*

11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*

12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*

13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*

14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*

15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*

16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*

17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*

18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*

19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*

20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*

21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*

22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	20	21	AA03086
2	107	100.0	20	22	AA031574
3	107	100.0	223	20	AA081767
4	107	100.0	223	21	AA069973
5	107	100.0	224	15	AA053637
6	107	100.0	229	21	AA077494
7	107	100.0	230	15	AA053638
8	107	100.0	231	21	AA091926
9	107	100.0	231	22	AA080953
10	107	100.0	233	21	AA091925

11	98	91.6	223	23	AB083322	Partial trypsin se
12	96	89.7	247	18	AA008475	Porcine trypsinoge
13	92	86.0	20	21	AA078977	Canine anionic try
14	92	86.0	246	21	AA078975	Canine anionic try
15	87	81.3	240	19	AA057740	Trypsinogen-like p
16	87	81.3	247	9	AA081243	Human spleen tryps
17	87	81.3	247	16	AA082703	Human pancreatic t
18	87	81.3	247	21	AA078974	Canine cationic tr
19	86	80.4	247	21	AA035701	Human trypsin hL a
20	85	79.4	224	20	AA031160	Human trypsin seri
21	85	79.4	225	22	AA098503	Human trypsin seri
22	85	79.4	230	20	AA093488	Human TRYI trypsin
23	85	79.4	241	21	AA021316	Human trypsinogen
24	85	79.4	246	19	AA064260	Human amyloid beta
25	85	79.4	247	21	AA021321	Human trypsinogen
26	85	79.4	254	23	AA074758	Human protease PRT
27	84	78.5	247	23	AA080693	Human pancreatic t
28	67	62.6	25	20	AA033346	Atlantic cod hydro
29	67	62.6	25	21	AA022965	Salmon enzyme 1 mu
30	67	62.6	25	21	AA093936	N-terminal of an a
31	67	62.6	25	22	AA007942	N-terminal of slam
32	67	62.6	25	23	AA021357	Salmon enzyme 1 mu
33	67	62.6	240	23	AB004644	Engraulis japonicu
34	65	60.7	37	21	AA093946	N-terminal of an a
35	65	60.7	37	23	AA021360	Atlantic cod multi
36	65	60.7	241	23	AB004645	Engraulis japonicu
37	61	57.0	225	22	AA031579	Amino acid sequenc
38	60	56.1	23	21	AA093937	N-terminal of an a
39	60	56.1	23	22	AA007943	N-terminal of atla
40	60	56.1	23	22	AA007944	N-terminal of atla
41	60	56.1	25	20	AA033347	Atlantic cod hydro
42	60	56.1	25	20	AA033348	Kill-derived multi
43	60	56.1	25	21	AA022966	Atlantic cod multi
44	60	56.1	25	21	AA022967	Atlantic cod multi
45	60	56.1	25	23	AA021358	Atlantic cod 1 mul

#### ALIGNMENTS

RESULT 1

AA03086

ID AA03086 standard; peptide; 20 AA.

XX AA03086;

XX 10-OCT-2000 (first entry)

XX N-terminus of bovine trypsin.

DE Trypsin; bovine; serine protease; Trichoderma; coagulant; hypertensive;

KW antinflammatory; leather preparation; silk treatment.

XX Bos taurus.

XX JP2000116377-A.

XX 25-APR-2000.

XX 08-OCT-1998; 98JP-0303263.

XX 08-OCT-1998; 98JP-0303263.

PR (AMAN ) AMANO PHARM KK.

PA WPI; 2000-369402/32.

XX A new serine protease and its preparation, used clinically in blood

PT coagulation, hypotension and anti-inflammation -

XX Example 3; Page 7; 9pp; Japanese.

PS The invention relates to a novel serine protease from Trichoderma sp.

CC

CC No.9064. The N-terminus of this protease is given in AAB03085. The novel  
 CC protease has trypsin-like activity, specifically cleaving peptide chains  
 CC on the carboxyl side of a basic amino acid (e.g., arginine or lysine).  
 CC It has a pH optimum of 7-8, and is stable between pH 5 and pH10. It has  
 CC a temperature optimum of approximately 40 degrees Celsius, and is stable  
 CC between 40 and 50 degrees Celsius. The invention also relates to a  
 CC method for the preparation of the novel serine protease, and the use of  
 CC the protease in protein degradation. The enzyme has coagulant,  
 CC hypertensive and anti-inflammatory effects. It may also be used in the  
 CC preparation of leather, for raw silk treatment and for the preparation of  
 CC protein hydrolysate. Sequences AAB03086-B03089 represent the N-terminal  
 CC 20 amino acids of trypsin from a variety of organisms.

SQ Sequence 20 AA;  
 Query Match 100.0%; Score 107; DB 21; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20  
 |||||  
 Db 1 IVGGYTCGANTVPYQVSLNS 20

RESULT 2  
 AAB31574  
 ID AAB31574 standard; peptide; 20 AA.

XX AAB31574;

DT 20-APR-2001 (first entry)

XX Amino terminal sequence of bovine trypsins.

DE Fish; serine proteinase; pain; acute inflammation; chronic inflammation;  
 KW arthritis; inflamed joint; bursitis; osteoarthritis; septic arthritis;  
 KW rheumatoid arthritis; juvenile rheumatoid arthritis; fibromyalgia;  
 KW systemic lupus erythematosus; phlebitis; tendinitis; rash; psoriasis;  
 KW acne; eczema; facial seboreic eczema; foreskin infection;  
 KW athlete's foot; fistulae infection; ulcer; navel infection; wrinkle;  
 KW scar; keloid; boil; wart; allergic itch; hemorrhoid; wound;  
 KW fungal infection; autoimmune disease.

XX Bos sp.

XX WO200078332-A2.

XX 28-DEC-2000.

XX 15-JUN-2000; 2000WO-1800005.

XX 18-JUN-1999; 99IS-0005086.

XX (BJAR/) BJARNASON J B.

XX Bjarnason JB;

XX WPI; 2001-091493/10.

XX Fish serine proteinase, useful as a cosmetic, medicament for treating  
 PT eczema, psoriasis, arthritis, and in the manufacture of the medicament  
 PT for treating, preventing pathogenic diseases involving receptor  
 PT mediated binding -

XX Disclosure; Page 5; 38pp; English.

XX The specification describes a fish serine proteinase. The proteinases  
 CC are useful as medicaments, for treating and preventing a disease in a  
 CC human or an animal such as pain, acute inflammation, chronic  
 CC inflammation, arthritis, inflamed joints, bursitis, osteoarthritis,  
 CC rheumatoid arthritis, juvenile rheumatoid arthritis, septic arthritis,  
 CC fibromyalgia, systemic lupus erythematosus, phlebitis, tendinitis, rash,  
 CC psoriasis, acne, eczema, facial seboreic eczema, eczema of the hands,

CC face or neck, foreskin infections, athlete's foot, fistulae infections,  
 CC infected topical ulcers, navel infections in newborns, wrinkles, scars,  
 CC keloids, boils, warts and allergic itch, hemorrhoids, wounds, wound  
 CC infections, wounds from burns, fungal infections and immunological and  
 CC autoimmune diseases. They are also useful for removing dead or peeling  
 CC skin from otherwise healthy skin, and for treating or preventing a  
 CC disease in which pathogenesis is caused by bacteria, virus, fungus,  
 CC parasite or a protozoan or a receptor mediated binding is involved.  
 CC The present sequence represents the amino terminal of bovine trypsins.

SQ Sequence 20 AA;

Query Match 100.0%; Score 107; DB 22; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20  
 |||||  
 Db 1 IVGGYTCGANTVPYQVSLNS 20

RESULT 3  
 AAW81767  
 ID AAW81767 standard; peptide; 223 AA.

XX AAW81767;

DT 29-JAN-1999 (first entry)

XX Bovine TRYP peptide fragment.

XX DHFR; dihydrofolic acid reductase; protein function; trypsin; bovine;  
 KW ribonuclease; myoglobin; database; homology; resemblance.

XX Bos taurus.

XX JF10287696-A.

XX 27-OCT-1998.

XX 11-APR-1997; 97JP-0093577.

XX 11-APR-1997; 97JP-0093577.

XX (IYAK-) IYAKU BUNSHI SEKKEI KENKYUSHO KK.

XX WPI; 1999-018384/02.

XX Determination of protein biological function - comprises use of  
 PT amino acid sequences databases containing the relevant information

XX Example 1; Fig 1; 11pp; Japanese.

XX This sequence is used in the creation of a database containing the  
 CC information for amino acid sequence of protein with at least 1  
 CC biological function with added a score on importance of expression of  
 CC the biological information for each amino acid residue. The database is  
 CC useful for determination of unknown biological function of a protein or  
 CC polypeptide based on the homology of amino acid sequence, e.g. steric  
 CC structure of protein, and includes retrieval and evaluation of high  
 CC homologous relationship for the determination of mostly resembling  
 CC protein. The database allows for correct and rapid retrieval and  
 CC presumption of protein and polypeptide having biological functions.

SQ Sequence 223 AA;

Query Match 100.0%; Score 107; DB 20; Length 223;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20  
 |||||  
 Db 1 IVGGYTCGANTVPYQVSLNS 20

```

RESULT 4
AAV69973
ID AAV69973 standard; protein; 223 AA.
XX AC AAV69973;
XX DT 13-APR-2000 (first entry)
XX DE TRYP protein.
XX KW Protein sequence database; biological function determination;
XX KW enzymatic activity; signaling activity; protein function determination;
XX KW TRYP protein.
XX OS Bos sp.
XX PN WO9962004-A1.
XX PD 02-DEC-1999.
XX PF 26-MAY-1998; 98WO-JP02302.
XX PR 26-MAY-1998; 98WO-JP02302.
XX PA (MEDI-) INST MEDICINAL MOLECULAR DESIGN INC.
XX PI Itai A, Tomioka N, Itai R, Imamura M;
XX DR WPI; 2000-136797/12.
XX KW An efficient, accurate and rapid computer database for estimating
PT protein functions e.g. enzymatic activity, for polypeptides obtained
PT from gene sequence translation -
XX PS Example 1; Fig 1; 26pp; Japanese.
XX CC This sequence represents the TRYP protein.
CC The invention relates to a database containing information on the amino
CC acid (aa) sequences of proteins of which 1 or more biological functions
CC are known. The database also contains additional information on the score
CC of importance of each aa residue in the whole aa sequence in determining
CC the known biological functions. The invention also relates to a method of
CC preparing an alignment between aa sequences contained in the database and
CC those of the unknown polypeptide. This is represented as the homology
CC amongst various sites, each being identified as having a high score of
CC importance in determining potential biological functions. The method is
CC used to enable an efficient estimation of the biological functions
CC (particularly enzymatic and signaling activities) of polypeptides from
CC their aa sequences. Suitable proteins can then be isolated and purified
CC by various means. This could be of considerable use in a biological and
CC medical context. The computerised procedure is efficient, fast and
CC accurate.
XX SQ Sequence 223 AA;
Query Match 100.0%; Score 107; DB 21; Length 223;
Best Local Similarity 100.0%; Pred. No. 2.9e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 IVGGYTCGANTVPYQVSLS 20
DB 1 IVGGYTCGANTVPYQVSLS 20
|||||
RESULT 5
AAR53637
ID AAR53637 standard; Protein; 224 AA.
XX AC AAR53637;
XX DT 03-JAN-1995 (first entry)

```

```

XX DE Bovine trypsin.
XX KW Cattle; cow; trypsin; enzyme; protease; proinsulin; insulin;
XX KW hormone; plasmid pRMG4.
XX OS Bos taurus.
XX PN EP597681-A.
XX PD 18-MAY-1994.
XX PF 10-NOV-1993; 93EP-0308959.
XX PR 13-NOV-1992; 92US-0977703.
XX PA (ELIL ) LILLY & CO ELI.
XX PI Greaney MG, Rosteck PR;
XX DR WPI; 1994-160671/20.
XX DR N-PSDB; AAQ63794.
XX PT Expression vectors for bovine trypsin and bovine trypsinogen -
PT for cleavage of zymogens into active drugs, e.g. pro-insulin
PT conversion into insulin
XX PS Disclosure; Page 25; 35pp; English.
XX CC This protein may be expressed by E. coli/plasmid pRMG4 and is able
CC to cleave zymogens into active drugs, e.g. pro-insulin conversion
CC into insulin.
XX SQ Sequence 224 AA;
Query Match 100.0%; Score 107; DB 15; Length 224;
Best Local Similarity 100.0%; Pred. No. 2.9e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 IVGGYTCGANTVPYQVSLS 20
DB 2 IVGGYTCGANTVPYQVSLS 21
|||||
RESULT 6
AAV77494
ID AAV77494 standard; Protein; 229 AA.
XX AC AAV77494;
XX DT 05-JUN-2000 (first entry)
XX DE Bovine trypsinogen.
XX KW Bovine; trypsinogen; heterologous protease; expression; transgenic plant;
XX KW large-scale production.
XX OS Bos taurus.
XX PH Key Location/Qualifiers
XX FT Protein 7..229 /note= "Bovine trypsinogen"
XX FT WO200005384-A1.
XX PN 03-FEB-2000.
XX PD 20-JUL-1999; 99WO-US12835.
XX PR 22-JUL-1998; 98US-0120582.
XX PA (PROD-) PRODIGENE INC.
XX

```

PI Howard JA, Hood E;  
 XX  
 DR WPI; 2000-195105/17.  
 DR N-PSDB; AA292585.  
 XX  
 PT Transgenic plant containing DNA encoding a heterologous protein, used  
 PT for the commercial production of protease -  
 XX  
 PS Example 1; Page 28-29; 32pp; English.  
 XX  
 CC The invention relates to a transgenic plant that comprises DNA encoding  
 CC an active or inactive form of a heterologous protease (e.g., trypsin)  
 CC The nucleotide sequence encoding the protease is operably linked to a  
 CC promoter to effect expression of the protease in the plant. The  
 CC transgenic plant is used in a method of producing protease in commercial  
 CC quantities, comprising the extraction of the protease from the plant  
 CC biomass. The current source of trypsin is the organs of animals. One  
 CC of the main difficulties with this source is that there is considerable  
 CC contamination by other proteases. Chymotrypsin is one of the additional  
 CC in the contaminants, and may cleave the product in an undesired manner.  
 CC The method of the invention overcomes the problem of contamination by  
 CC other additional proteases. By directing the expression of the protease  
 CC to the seed of the plant, the problem of expressing the protease and  
 CC having it digest the proteins in the plant, resulting in cell death is  
 CC overcome. The present sequence represents bovine trypsinogen. cDNA  
 CC (Genbank accession #00760) encoding this protease was used in the  
 CC construction of a plant expression plasmid in the exemplifications of  
 CC the invention.  
 XX

Sequence 229 AA;  
 Query Match 100.0%; Score 107; DB 21; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 3e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IVGGYTCGANTVPYQVSLNS 20  
 |||||  
 Db 7 IVGGYTCGANTVPYQVSLNS 26  
 |||||

RESULT 7  
 AAR53638  
 ID AAR53638 standard; Protein; 230 AA.  
 XX  
 AC AAR53638;  
 XX  
 DT 01-DEC-1994 (first entry)  
 XX  
 DE Bovine trypsinogen.  
 XX  
 KW Cattle; cow; trypsinogen; enzyme; protease; proinsulin; insulin;  
 KW hormone; plasmid pRMG4.  
 XX  
 OS Bos taurus.  
 XX  
 PN EP597681-A.  
 XX  
 PD 18-MAY-1994.  
 XX  
 PF 10-NOV-1993; 93EP-0308959.  
 XX  
 PR 13-NOV-1992; 92US-0977703.  
 XX  
 PA (ELIL ) LILLY & CO ELI.  
 XX  
 PI Greaney MG, Rosteck PR;  
 XX  
 DR WPI; 1994-160671/20.  
 XX  
 PT Expression vectors for bovine trypsin and bovine trypsinogen -  
 PT for cleavage of zymogens into active drugs, e.g. pro-insulin  
 PT conversion into insulin  
 XX

PS Disclosure; Page 27; 35pp; English.  
 XX  
 CC This protein may be expressed by E. coli/plasmid pRMG7 and is able  
 CC to cleave zymogens into active drugs, e.g. pro-insulin conversion  
 CC into insulin.  
 XX

SQ Sequence 230 AA;

Query Match 100.0%; Score 107; DB 15; Length 230;  
 Best Local Similarity 100.0%; Pred. No. 3e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 IVGGYTCGANTVPYQVSLNS 20  
 |||||  
 Db 8 IVGGYTCGANTVPYQVSLNS 27  
 |||||

RESULT 8  
 AAY91926  
 ID AAY91926 standard; Protein; 231 AA.  
 XX  
 AC AAY91926;  
 XX  
 DT 19-JUL-2000 (first entry)  
 XX  
 DE Recombinant trypsin.  
 XX  
 KW Recombinant trypsin; trypsinogen analogue; mutated bovine trypsinogen;  
 KW leader sequence; trypsin activity; recombinant protein production;  
 KW inactive zymogen.  
 XX  
 OS Synthetic.  
 OS Bos taurus.  
 XX

FH Key Location/Qualifiers  
 FT Peptide 1..2  
 FT Peptide /note= "optionally absent, see disclosure"  
 FT Peptide 3..7  
 FT /label= leader\_sequence  
 FT Protein 8..231  
 FT /label= mature\_protein  
 XX  
 PN WO200017332-A1.  
 XX  
 PD 30-MAR-2000.  
 XX

PF 15-SEP-1999; 99WO-US21047.  
 XX

PR 21-SEP-1998; 98US-0101213.  
 XX

PA (ELIL ) LILLY & CO ELI.  
 XX

PI Hanquier JM, Hershberger CL, Desplancq D, Larson JL, Rosteck PR;  
 XX

DR WPI; 2000-283565/24.  
 XX

DR N-PSDB; AAA08526.  
 XX

PT New trypsinogen analog useful for the production of recombinant trypsin  
 PT has a modified leader sequence not cleavable by trypsin or trypsin-like  
 PT enzymes

PS Claim 22; Page 51-52; 56pp; English.  
 XX

CC This sequence shows a claimed recombinant trypsin. The trypsin is  
 CC produced by cleavage of a trypsinogen analogue (AAY91925). A wild type  
 CC bovine trypsinogen was mutated to destroy the trypsin cleavage site.  
 CC The lys residue present in the leader sequence of the native bovine  
 CC trypsinogen protein was mutated to an Asp residue. The vector was  
 CC constructed such that DNA encoding a (Glu-Ala)2 peptide was inserted  
 CC between the C-terminus of the alpha factor signal and the N-terminus of  
 CC the trypsinogen analogue leader sequence to improve the yield of the  
 CC secreted protein. The specification claims an isolated trypsinogen  
 CC analogue comprising a protein having trypsin activity and a leader



CC sequence having at least two amino acids which are not Lys or Arg.  
 CC The trypsin derived from the recombinant trypsinogen is useful for the  
 CC characterization of other proteins, and in the manufacture of other  
 CC recombinant bioproducts, for example to cleave leader sequences  
 CC from small recombinant proteins expressed initially as fusion proteins.  
 CC The present method provides for expression of an inactive zymogen form  
 CC that is soluble and properly folded yet is not activated until after  
 CC purification from fermentation broth or cell extracts. This is  
 CC accomplished through the expression of a single chain trypsinogen  
 CC analogue where the leader sequence is modified such that it lacks a  
 CC trypsin-like enzyme cleavage site. Specifically the trypsinogen  
 CC analogues of the present invention lack a lysin or arginine in the  
 CC N-terminal leader sequence of the protein to prevent auto-activation or  
 CC activation by endogenous host cell enzymes.

XX XX Sequence 231 AA;  
 Query Match 100.0%; Score 107; DB 21; Length 231;  
 Best Local Similarity 100.0%; Pred. No. 3e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20  
 Db 9 IVGGYTCGANTVPYQVSLNS 28  
 |||||

RESULT 9  
 AAB80953  
 ID AAB80953 standard; Protein; 231 AA.  
 XX AAB80953;  
 AC AAB80953;  
 DT 06-JUN-2001 (first entry)  
 DE Bovine met-phe-trypsinogen.  
 XX Trypsinogen; bovine; trypsin; serine protease.  
 OS Bos sp.  
 XX Key Location/Qualifiers  
 FH Peptide 4..8  
 FT /note= "Leader peptide, cleaved to produce trypsin"  
 FT WO200119970-A2.  
 XX 22-MAR-2001.  
 XX 05-SEP-2000; 2000WO-US20813.  
 XX 15-SEP-1999; 99US-0154019.  
 XX (ELIL ) LILLY & CO ELI.  
 XX Hanquier JM, Hershberger CL, Larson JL, Rosteck PR;  
 XX WPI: 2001-273425/28.  
 XX N-PSDB; AAF81479.  
 XX New chymotrypsin-free trypsin and trypsinogen useful for manufacturing  
 FT recombinant protein pharmaceuticals and pure trypsin -  
 XX Claim 20; Fig 1; 55pp; English.

CC The present sequence is bovine met-phe-trypsinogen. Trypsin is a serine  
 CC protease which cleaves the peptide bond on the carboxy-terminus of basic  
 CC amino acid residues. Trypsin is synthesised in a slightly longer  
 CC catalytically inactive form: trypsinogen, which itself is cleaved (leader  
 CC sequence removed) to give trypsin. The leader sequence of the present  
 CC protein consists of (Asp)4-Lys and is present at the amino-terminus.  
 CC This protein has two additional residues at the amino terminus: Met and  
 CC Phe. Bovine met-phe-trypsinogen is useful for the manufacture of  
 CC recombinant protein pharmaceuticals. High purity trypsin products are

CC produced by expressing trypsinogen inside a prokaryotic cell which is  
 CC then isolated and activated to form trypsin.

XX XX Sequence 231 AA;  
 Query Match 100.0%; Score 107; DB 22; Length 231;  
 Best Local Similarity 100.0%; Pred. No. 3e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20  
 Db 9 IVGGYTCGANTVPYQVSLNS 28  
 |||||

RESULT 10  
 AAY91925  
 ID AAY91925 standard; Protein; 233 AA.  
 XX AAY91925;  
 AC AAY91925;  
 DT 19-JUL-2000 (first entry)  
 DE Trypsinogen analogue.  
 XX Trypsinogen analogue; mutated bovine trypsinogen; leader sequence;  
 KW trypsin activity; recombinant protein production; inactive zymogen.  
 XX Synthetic.  
 OS Bos taurus.  
 XX Key Location/Qualifiers  
 FH Peptide 1..3  
 FT /note= "optionally absent, see disclosure"  
 FT Peptide 1..4  
 FT /label= linker\_peptide  
 FT /note= "increases yield of secreted protein"  
 FT 5..10  
 FT /label= leader\_sequence  
 FT Misc-difference 10  
 FT /note= "native bovine trypsinogen protein was  
 FT mutated to an Asp residue"  
 FT Protein 11..233  
 FT /label= mature protein  
 FT /note= "trypsin"  
 FT Modified-site 169  
 FT /note= "potential N-glycosylation"  
 XX WO200017332-A1.  
 XX 30-MAR-2000.  
 XX 15-SEP-1999; 99WO-US21047.  
 XX 21-SEP-1998; 98US-0101213.  
 XX (ELIL ) LILLY & CO ELI.  
 XX Hanquier JM, Hershberger CL, Desplancq D, Larson JL, Rosteck PR;  
 XX WPI: 2000-283565/24.  
 XX N-PSDB; AAA08525.  
 XX New trypsinogen analog useful for the production of recombinant trypsin  
 PT has a modified leader sequence not cleavable by trypsin or trypsin-like  
 PT enzymes  
 XX Claim 5, 6; Page 47-48; 56pp; English.

CC This sequence shows a trypsinogen analogue. The wild type bovine  
 CC trypsinogen was mutated to destroy the trypsin cleavage site. The lys  
 CC residue present in the leader sequence of the native bovine trypsinogen  
 CC protein was mutated to an Asp residue. The vector was constructed such  
 CC that DNA encoding a (Glu-Ala)2 peptide was inserted between the

CC C-terminus of the alpha factor signal and the N-terminus of the  
 CC trypsinogen analogue leader sequence to improve the yield of the  
 CC secreted protein. The specification claims an isolated trypsinogen  
 CC analogue comprising a protein having trypsin activity and a leader  
 CC sequence having at least two amino acids which are not lys or Arg. A  
 CC recombinantly produced trypsin (AA91926) is also claimed. The trypsin  
 CC derived from the recombinant trypsinogen is useful for the  
 CC characterization of other proteins, and in the manufacture of other  
 CC recombinant bioproducts, for example to cleave leader sequences  
 CC from small recombinant proteins expressed initially as fusion proteins.  
 CC The present method provides for expression of an inactive zymogen form  
 CC that is soluble and properly folded yet is not activated until after  
 CC purification from fermentation broth or cell extracts. This is  
 CC accomplished through the expression of a single chain trypsinogen  
 CC analogue where the leader sequence is modified such that it lacks a  
 CC trypsin-like enzyme cleavage site. Specifically the trypsinogen  
 CC analogues of the present invention lack a lysin or arginine in the  
 CC N-terminal leader sequence of the protein to prevent auto-activation or  
 CC activation by endogenous host cell enzymes.

XX Sequence 233 AA;

Query Match 100.0%; Score 107; DB 21; Length 233;  
 Best Local Similarity 100.0%; Pred. No. 3e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IVGGYTCGANTVPYQVSLNS 20  
 Db 11 IVGGYTCGANTVPYQVSLNS 30

RESULT 11

AB83322  
 ID AB83322 standard; protein; 223 AA.

XX ABB83322;

DT 30-AUG-2002 (first entry)

XX Partial trypsin sequence.

XX Trypsin; three dimensional structure; protein structure analysis;  
 KW serine protease.

XX Unidentified.

XX US2002035434-A1.

XX 21-MAR-2002.

XX 23-JUL-2001; 2001US-0910054.

XX 06-FEB-1992; 92JP-0021012.

XX 11-DEC-1992; 92JP-0331703.

XX 08-FEB-1993; 93US-0014867.

XX (FUJITSU) FUJITSU LTD.

XX Tomikawa M, Aikawa S, Matsuzawa F;

XX WPI; 2002-507172/54.

XX Analysis of three-dimensional structures by generating combination of  
 PT correspondence satisfying restriction condition, and calculating root  
 PT mean square distance between elements in the combination of  
 PT correspondence

XX Disclosure; Fig 38; 65pp; English.

XX The present invention relates to a method for analysis of three  
 CC dimensional structures. The method involves generating a combination of  
 CC correspondence satisfying a restriction condition between the elements  
 CC belonging to a first and second point sets from among all candidates for

CC the combination of correspondence, and calculating a root mean square  
 CC distance between the elements corresponding in the combination of  
 CC correspondence. The method is useful for analysing three-dimensional  
 CC structures of molecules, particularly proteins. To illustrate the  
 CC invention, the serine proteases trypsin and elastase were used. The  
 CC present sequence is a partial sequence for trypsin. The histidine and  
 CC serine active sites of trypsin and elastase were analysed using the  
 CC method of the invention.

XX Sequence 223 AA;

Query Match 91.6%; Score 98; DB 23; Length 223;  
 Best Local Similarity 95.0%; Pred. No. 6.6e-07;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IVGGYTCGANTVPYQVSLNS 20  
 Db 1 IVGGYTCGANTVPYQVSLNS 20

RESULT 12

AAW08475  
 ID AAW08475 standard; Protein; 247 AA.

XX AAW08475;

DT 28-MAR-1997 (first entry)

XX Porcine trypsinogen.

XX Trypsinogen; trypsin; protease; filamentous fungus; Aspergillus.

XX Sus scrofa.

XX WO9700316-A1.

XX 03-JAN-1997.

XX 10-JUN-1996; 96WO-DK00253.

XX 16-JUN-1995; 95DK-0000693.

XX (NOVO) NOVO-NORDISK AS.

XX Kjeldsen TB, Woldike HF;

XX WPI; 1997-077523/07.

XX N-PSDB; AAT49878.

XX Prodn. of trypsin or trypsinogen in filamentous fungi - transformed  
 PT with recombinant DNA comprising trypsinogen sequence fused to signal  
 PT sequence, provides higher level of expression than other expression  
 PT systems

XX Example 2; Page 15; 28pp; English.

XX An almost complete pig trypsinogen (AAW08475) lacking the very N-  
 CC terminal end of the signal peptide is the product of a cDNA clone  
 CC (AAT49878) isolated from a pig pancreas cDNA library. The cDNA clone  
 CC can be fused to a signal sequence (e.g. from the native enzyme or  
 CC from a specified amylose gene) and used for prodn. of recombinant  
 CC trypsin in transformed host cells, esp. filamentous fungus, e.g.  
 CC Aspergillus oryzae or Aspergillus niger. Levels of trypsin  
 CC secreted by Aspergillus spp. are several-fold higher than those  
 CC obtd. using other microbial systems.

XX Sequence 247 AA;

Query Match 89.7%; Score 96; DB 18; Length 247;  
 Best Local Similarity 85.0%; Pred. No. 1.5e-06;  
 Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IVGGYTCGANTVPYQVSLNS 20

```

Db      25 IVGGYTCAANSIPVQVSLNS 44
      ||||| ||::|||
RESULT 13
AAAY78977
ID AAY78977 standard; protein; 20 AA.
XX AC AAY78977;
XX DT 05-JUN-2000 (first entry)
XX DE Canine anionic trypsinogen immunogenetic fragment #2 amino acid sequence.
XX KW Anionic trypsinogen; dog; monoclonal antibody production; detect;
XX KW diagnose; pancreatitis; pancreatic cancer; renal insufficiency;
XX KW extrapancreatic hyposecretion; immunogenetic fragment.
XX OS Canis familiaris.
XX PN W0200009739-A1.
XX PD 24-FEB-2000.
XX PF 09-AUG-1999; 99WO-JP04299.
XX PR 10-AUG-1998; 98JP-0236609.
XX PR 10-MAR-1999; 99JP-0063990.
XX PA (FUJY ) FUJI YAKUHIN KOGYO KK.
XX PI Waritani T, Ashida Y, Yamada T;
XX PI WPI; 2000-206018/18.
XX PT Anti-canine trypsin monoclonal antibody, useful as reagent for quick
XX PT and accurate detection and quantitation of trypsin and/or trypsin-like
XX PT immunoreactants in various forms in diagnosis e.g. of pancreatic
XX PT diseases -
XX PS Claim 3; Page 64-65; 67pp; Japanese.
XX CC This sequence represents an immunogenetic fragment of the canine anionic
XX CC trypsinogen amino acid sequence. The fragment corresponds to amino acid
XX CC residues 24 to 43 of the anionic trypsinogen protein sequence
XX CC (see AAY78975). The invention relates to monoclonal antibodies with
XX CC specificity against canine trypsin, or canine trypsin-related
XX CC substances. The antibodies are highly specific and can be used as reagent
XX CC for quick and accurate detection and quantitation of canine trypsin and
XX CC canine trypsin-like immunoreactants in various forms. The antibodies can
XX CC be used in the diagnosis of diseases such as pancreatitis, pancreatic
XX CC cancer, renal insufficiency and extrapancreatic hyposecretion.
XX SQ Sequence 20 AA;
      Query Match 86.0%; Score 92; DB 21; Length 20;
      Best Local Similarity 85.0%; Pred. No. 3.6e-07;
      Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY 1 IVGGYTCGANTVPYQVSLNS 20
      ||||| ||::|||
Db 1 IVGGYTCRSNPVQVSLNS 20
      ||||| ||::|||
RESULT 14
AAAY78975
ID AAY78975 standard; protein; 246 AA.
XX AC AAY78975;
XX DT 05-JUN-2000 (first entry)
XX DE Canine anionic trypsinogen amino acid sequence.

```

```

XX Anionic trypsinogen; dog; monoclonal antibody production; detect;
XX KW diagnose; pancreatitis; pancreatic cancer; renal insufficiency;
XX KW extrapancreatic hyposecretion.
XX OS Canis familiaris.
XX PN W0200009739-A1.
XX PD 24-FEB-2000.
XX PF 09-AUG-1999; 99WO-JP04299.
XX PR 10-AUG-1998; 98JP-0236609.
XX PR 10-MAR-1999; 99JP-0063990.
XX PA (FUJY ) FUJI YAKUHIN KOGYO KK.
XX PI Waritani T, Ashida Y, Yamada T;
XX PI WPI; 2000-206018/18.
XX PT Anti-canine trypsin monoclonal antibody, useful as reagent for quick
XX PT and accurate detection and quantitation of trypsin and/or trypsin-like
XX PT immunoreactants in various forms in diagnosis e.g. of pancreatic
XX PT diseases -
XX PS Claim 3; Page 64-65; 67pp; Japanese.
XX CC This sequence represents the canine anionic trypsinogen amino acid
XX CC sequence. The protein was isolated from the canine pancreas. The
XX CC invention relates to monoclonal antibodies with specificity against
XX CC canine trypsin, or canine trypsin-related substances. The antibodies are
XX CC highly specific and can be used as reagent for quick and accurate
XX CC detection and quantitation of canine trypsin and canine trypsin-like
XX CC immunoreactants in various forms. The antibodies can be used in the
XX CC diagnosis of diseases such as pancreatitis, pancreatic cancer, renal
XX CC insufficiency and extrapancreatic hyposecretion.
XX SQ Sequence 246 AA;
      Query Match 86.0%; Score 92; DB 21; Length 246;
      Best Local Similarity 85.0%; Pred. No. 6e-06;
      Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY 1 IVGGYTCGANTVPYQVSLNS 20
      ||||| ||::|||
Db 24 IVGGYTCRSNPVQVSLNS 43
      ||||| ||::|||
RESULT 15
AAW57740
ID AAW57740 standard; Protein; 240 AA.
XX AC AAW57740;
XX DT 16-SEP-1998 (first entry)
XX DE Trypsinogen-like protein.
XX KW Trypsinogen-like protein; human.
XX OS Homo sapiens.
XX PN JP10099080-A.
XX PD 21-APR-1998.
XX PR 26-SEP-1996; 96JP-0273923.
XX PR 26-SEP-1996; 96JP-0273923.
XX PA (SHIS ) SHISEIDO CO LTD.

```

```

XX  WPI; 1998-289873/26.
DR  N-PSDB; AAV24548.
XX
XX  DNA coding for trypsinogen-like protein
XX
XX  Claim 3; Page 4-5; 7pp; Japanese.
XX
XX  This sequence represents a human trypsinogen-like protein, and is encoded
CC  by the DNA of the invention.
XX
XX  SQ Sequence 240 AA;
XX
XX  Query Match      81.3%; Score 87; DB 19; Length 240;
XX  Best Local Similarity 80.0%; Pred. No. 3.3e-05;
XX  Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
XX
Qy  1 IVGGYTCGANTVPYQVSLNS 20
Db  17 IVGGYTCBENSLLPYQVSLNS 36

```

Search completed: February 12, 2003, 10:22:22  
Job time : 15.4328 secs

[illegible]

20	52	54.7	262	10	US-09-988-975A-7
21	52	54.7	267	10	US-09-925-297-745
22	52	54.7	375	10	US-09-755-100-11
23	52	54.7	449	9	US-10-012-896-617
24	52	54.7	449	9	US-09-895-793-617
25	52	54.7	449	9	US-09-895-614-617
26	52	54.7	449	10	US-09-159-143-617
27	52	54.7	449	10	US-09-780-669-617
28	52	54.7	449	10	US-09-822-827-617
29	52	54.7	585	10	US-09-895-793-978
30	52	54.7	801	9	US-09-832-827-978
31	52	54.7	801	9	US-09-895-793-982
32	52	54.7	801	10	US-09-832-827-982
33	52	54.7	1079	10	US-09-895-793-947
34	52	54.7	1079	10	US-09-822-827-947
35	47	49.5	260	9	US-09-978-295A-395
36	47	49.5	260	9	US-09-978-697-395
37	47	49.5	260	9	US-09-978-192A-395
38	47	49.5	260	9	US-09-999-832A-395
39	47	49.5	260	9	US-09-978-189-395
40	47	49.5	260	9	US-10-028-072-396
41	47	49.5	260	9	US-10-121-049-396
42	47	49.5	260	9	US-10-123-904-396
43	47	49.5	260	9	US-10-140-470-396
44	47	49.5	260	9	US-10-175-746-396
45	47	49.5	260	9	US-10-179-918-396

## ALIGNMENTS

```

RESULT 1
US-09-938-269-15
/ Sequence 15, Application US/09938269
/ Publication NO. US20030067951A1
/ GENERAL INFORMATION:
/ APPLICANT: Franklin, Richard L.
/ APPLICANT: Cowling, Didier S.P.
/ APPLICANT: Hubbel, Jeffrey A.
/ APPLICANT: van de Wetering, Petra
/ TITLE OF INVENTION: Treatment Of Trauma
/ FILE REFERENCE: 314572-105
/ CURRENT APPLICATION NUMBER: US/09/938,269
/ CURRENT FILING DATE: 2001-08-23
/ NUMBER OF SEQ ID NOS: 17
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 15
/ LENGTH: 25
/ TYPE: PRT
/ ORGANISM: Atlantic cod
US-09-938-269-15

```

```

Query Match      94.7%; Score 90; DB 9; Length 25;
Best Local Similarity 85.0%; Pred.No. 9.8e-09;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 IVGGYCXHXHQAQHVSLNS 20
        ||||| ||||| ||||| |||||
Db      1 IVGGYECTKHQAQHVSLNS 20

RESULT 2
US-09-938-269-16
; Sequence 16. Application US/09938269
; Publication No. US20030007951A1
; GENERAL INFORMATION:
; APPLICANT: Franklin, Richard L.
; APPLICANT: Cowling, Didier S.P.
; APPLICANT: Hubbel, Jeffrey A.
; APPLICANT: van de Watering, Petra
; TITLE OF INVENTION: Treatment of Trauma
; FILE REFERENCE: 314572-105
; CURRENT APPLICATION NUMBER: US/09/938,269

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:22:36 ; Search time 3.04478 Seconds  
(without alignments)  
167.821 Million cell updates/sec

Title: US-10-036-371-1

Perfect score: 95  
Sequence: 1 IVGGYXCXXHSHQAHQVSLNS 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seg length: 0

Minimum DB seq	length: 2000000000
Maximum DB seq	length: 2000000000

SECRET FOR THE INFORMATION

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications\_AA.\*

- 1: /cgn2\_6/prodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/prodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/prodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/prodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/prodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/prodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/prodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/prodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/prodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/prodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/prodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/prodata/1/pubpaa/US60\_PUBCOMB.pep.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	90	94.7	25	9	US-09-938-369-15	Sequence 15, Appl
2	90	94.7	25	9	US-09-938-369-16	Sequence 16, Appl
3	90	94.7	25	9	US-09-938-369-17	Sequence 17, Appl
4	90	94.7	225	12	US-10-036-371-8	Sequence 8, Appl
5	89	93.7	20	12	US-10-036-371-1	Sequence 1, Appl
6	78	82.1	25	9	US-09-938-369-14	Sequence 14, Appl
7	66	69.5	271	12	US-10-021-368-10	Sequence 10, Appl
8	64	67.4	247	10	US-09-923-779-154	Sequence 154, Appl
9	63	66.3	20	12	US-10-036-371-2	Sequence 2, Appl
10	62	65.3	223	10	US-09-910-071-14	Sequence 14, Appl
11	55	57.3	235	10	US-09-925-300-1093	Sequence 1093, Ap
12	55	57.9	286	12	US-10-021-368-9	Sequence 9, Appl
13	52	54.7	261	9	US-09-895-793-946	Sequence 946, App
14	52	54.7	261	9	US-10-131-241-62	Sequence 62, Appl
15	52	54.7	261	10	US-09-796-294-10	Sequence 10, Appl
16	52	54.7	261	10	US-09-822-827-946	Sequence 946, Appl
17	52	54.7	261	10	US-09-963-896-6	Sequence 6, Appl
18	52	54.7	261	10	US-09-767-403-1	Sequence 1, Appl
19	52	54.7	262	10	US-09-764-761-4	Sequence 4, Appl

```

; CURRENT FILING DATE: 2001-08-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Atlantic cod
US-09-938-269-16

```

```

Query Match      94.7%; Score 90; DB 9; Length 25;
Best Local Similarity 85.0%; Pred. No. 9.8e-09;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 1 IVGGYXCXHSQAHOVSLS 20
    ||||| ||||| |||||
Db 1 IVGGYECTRHQAHOVSLS 20

```

```

RESULT 3
US-09-938-269-17
; Sequence 17, Application US/09938269
; Publication No. US20030007951A1
; GENERAL INFORMATION:
; APPLICANT: Franklin, Richard L.
; APPLICANT: Cowling, Didier S.P.
; APPLICANT: Hubbel, Jeffrey A.
; APPLICANT: van de Wetering, Petra
; TITLE OF INVENTION: Treatment of Trauma
; FILE REFERENCE: 314572-105
; CURRENT APPLICATION NUMBER: US/09/938,269
; CURRENT FILING DATE: 2001-08-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Atlantic cod
US-09-938-269-17

```

```

Query Match      94.7%; Score 90; DB 9; Length 37;
Best Local Similarity 85.0%; Pred. No. 1.5e-08;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 1 IVGGYXCXHSQAHOVSLS 20
    ||||| ||||| |||||
Db 1 IVGGYQCEAHQAHOVSLS 20

```

```

RESULT 4
US-10-036-371-8
; Sequence 8, Application US/10036371
; Patent No. US20020141987A1
; GENERAL INFORMATION:
; APPLICANT: BJARNARSON, JON B.
; TITLE OF INVENTION: FISH SERINE PROTEINASES AND THEIR PHARMACEUTICAL AND
; FILE REFERENCE: 81691/284960
; CURRENT APPLICATION NUMBER: US/10/036,371
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: 09/411,688
; PRIOR FILING DATE: 1999-10-12
; PRIOR APPLICATION NUMBER: 5086/99
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Gadus sp.
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (9)_RES
; OTHER INFORMATION: K or R

```

```

; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (24)
; OTHER INFORMATION: Y or F
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (32)
; OTHER INFORMATION: K or E
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (33)
; OTHER INFORMATION: D or Q
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (64)
; OTHER INFORMATION: Y or F
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (71)
; OTHER INFORMATION: I or unknown
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (82)
; OTHER INFORMATION: N or D
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (92)
; OTHER INFORMATION: K or E
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (129)
; OTHER INFORMATION: A or D
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (136)
; OTHER INFORMATION: V or C
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (138)
; OTHER INFORMATION: N or S
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (204)
; OTHER INFORMATION: H or N
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (212)
; OTHER INFORMATION: M, V or C
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (224)
; OTHER INFORMATION: N or S
US-10-036-371-8

```

```

Query Match      94.7%; Score 90; DB 12; Length 225;
Best Local Similarity 90.0%; Pred. No. 9.3e-08;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 1 IVGGYXCXHSQAHOVSLS 20
    ||||| ||||| |||||
Db 1 IVGGYECTXHQAHOVSLS 20

```

```

RESULT 5
US-10-036-371-1
; Sequence 1, Application US/10036371
; Patent No. US20020141987A1
; GENERAL INFORMATION:
; APPLICANT: BJARNARSON, JON B.
; TITLE OF INVENTION: FISH SERINE PROTEINASES AND THEIR PHARMACEUTICAL AND
; TITLE OF INVENTION: COSMETIC USE
; FILE REFERENCE: 81691/284960
; CURRENT APPLICATION NUMBER: US/10/036,371

```

; CURRENT FILING DATE: 2002-01-07  
; PRIOR APPLICATION NUMBER: 09/411,688  
; PRIOR FILING DATE: 1999-10-12  
; PRIOR APPLICATION NUMBER: 5086/99  
; PRIOR FILING DATE: 1999-06-18  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Gadus sp.  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (6)  
; OTHER INFORMATION: Q or E  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (8)  
; OTHER INFORMATION: E or T  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (9)  
; OTHER INFORMATION: K or R  
; US-10-036-371-1

Query Match 93.7%; Score 89; DB 12; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.1e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVGGYXCXHSQAHOVSLS 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 IVGGYXCXHSQAHOVSLS 20

RESULT 6  
US-09-938-269-14  
; Sequence 14, Application US/09938269  
; Publication No. US20030007951A1  
; GENERAL INFORMATION:  
; APPLICANT: Franklin, Richard L.  
; APPLICANT: Cowling, Didier S.P.  
; APPLICANT: Hubbel, Jeffrey A.  
; APPLICANT: van de Metering, Petra  
; TITLE OF INVENTION: Treatment of Trauma  
; FILE REFERENCE: 314572-105  
; CURRENT APPLICATION NUMBER: US/09/938,269  
; CURRENT FILING DATE: 2001-08-23  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 25  
; TYPE: PRT  
; ORGANISM: Salmon enzyme  
; US-09-938-269-14

Query Match 82.1%; Score 78; DB 9; Length 25;  
Best Local Similarity 75.0%; Pred. No. 9.2e-07;  
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYXCXHSQAHOVSLS 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 IVGGYCKAYSQAHOVSLS 20

RESULT 7  
US-10-021-368-10  
; Sequence 10, Application US/10021368  
; Patent No. US20020106367A1  
; GENERAL INFORMATION:  
; APPLICANT: Band, Vimala  
; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
; MOLECULES AND METHODS  
; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/021,368  
; FILING DATE: 12-Dec-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/201,038  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00398/100002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 271 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
; US-10-021-368-10

Query Match 69.5%; Score 66; DB 12; Length 271;  
Best Local Similarity 53.6%; Pred. No. 0.00099;  
Matches 15; Conservative 1; Mismatches 4; Indels 8; Gaps 1;

QY 1 IVGGYX-----CXHSQAHOVSLS 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 10 IVGGYSTRYPIIECKAYSQPHOVSLS 37

RESULT 8  
US-09-923-779-154  
; Sequence 154, Application US/09923779  
; Patent No. US20020076721A1  
; GENERAL INFORMATION:  
; APPLICANT: Pyle, Ruth A.  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER  
; FILE REFERENCE: 210121.553  
; CURRENT APPLICATION NUMBER: US/09/923,779  
; CURRENT FILING DATE: 2001-08-06  
; NUMBER OF SEQ ID NOS: 155  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 154  
; LENGTH: 247  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-923-779-154

Query Match 67.4%; Score 64; DB 10; Length 247;  
Best Local Similarity 65.0%; Pred. No. 0.0019;  
Matches 13; Conservative 2; Mismatches 5; Indels 3; Gaps 0;

QY 1 IVGGYXCXHSQAHOVSLS 20  
| | | | | | | | | | | | | | | | | | | | | |

Db 24 IVGGYCEENSVPYQVSLNS 43

RESULT 9  
US-10-036-371-2  
; Sequence 2, Application US/10036371  
; Patent No. US20020141987A1  
; GENERAL INFORMATION:  
; APPLICANT: BJARNARSON, JON B.  
; TITLE OF INVENTION: FISH SERINE PROTEINASES AND THEIR PHARMACEUTICAL AND  
; FILE REFERENCE: 81691/284960  
; CURRENT APPLICATION NUMBER: US/10/036,371  
; CURRENT FILING DATE: 2002-01-07  
; PRIOR APPLICATION NUMBER: 09/411,688  
; PRIOR FILING DATE: 1999-10-12  
; PRIOR APPLICATION NUMBER: 5086/99  
; PRIOR FILING DATE: 1999-06-18  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Bovine sp.  
US-10-036-371-2

Query Match 66.3%; Score 63; DB 12; Length 20;  
Best Local Similarity 60.0%; Pred. No. 0.00021;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVGGYKXKXHSQAQVSLNS 20  
||||| : : : |||||  
Db 1 IVGGYTCGANTVPYQVSLNS 20

RESULT 10  
US-09-910-071-14  
; Sequence 14, Application US/09910071  
; Patent No. US20020116146A1  
; GENERAL INFORMATION:  
; APPLICANT: Tomikawa, Mayumi  
; APPLICANT: Aikawa, Seiichi  
; APPLICANT: Matsuzawa, Fumiko  
; TITLE OF INVENTION: Method and Apparatus for Extracting and Evaluating Mutually Similar  
; TITLE OF INVENTION: Portions in One-Dimensional Sequences in Molecules and/or Three-  
; TITLE OF INVENTION: Structures of Molecules  
; FILE REFERENCE: 522.1921D2  
; CURRENT APPLICATION NUMBER: US/09/910,071  
; CURRENT FILING DATE: 2001-07-23  
; PRIOR APPLICATION NUMBER: US 08/014,867  
; PRIOR FILING DATE: 1993-02-08  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 14  
; LENGTH: 223  
; TYPE: PRT  
; ORGANISM: unknown  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of trypsin  
US-09-910-071-14

Query Match 65.3%; Score 62; DB 10; Length 223;  
Best Local Similarity 60.0%; Pred. No. 0.0037;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVGGYKXKXHSQAQVSLNS 20  
||||| : : : |||||  
Db 1 IVGGYTCGANTVPYQVSLNS 20

RESULT 11  
US-09-925-300-1093  
; Sequence 1093, Application US/09925300

; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Ruben,  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA101  
; CURRENT APPLICATION NUMBER: US/09/925,300  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05988  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1090  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1093  
; LENGTH: 235  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-300-1093

Query Match 57.9%; Score 55; DB 10; Length 235;  
Best Local Similarity 55.0%; Pred. No. 0.055;  
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 IVGGYKXKXHSQAQVSLNS 20  
||||| : : : |||||  
Db 87 IVGGWCEKHSQWQVAVYS 106

RESULT 12  
US-10-021-368-9  
; Sequence 9, Application US/10021368  
; Patent No. US20020106367A1  
; GENERAL INFORMATION:  
; APPLICANT: Band, Vimla  
; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
; MOLECULES AND METHODS  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/021,368  
; FILING DATE: 12-Dec-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/201,038  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00398/100002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 286 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:





**THIS PAGE BLANK (USPTO)**



/ ADDRESSEE: Dechert Price & Rhoads  
/ STREET: 997 Lenox Drive, Building 3, Suite 210  
/ CITY: Lawrenceville  
/ STATE: NJ  
/ COUNTRY: USA  
/ ZIP: 08543  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Diskette  
/ COMPUTER: IBM Compatible  
/ OPERATING SYSTEM: Windows 95  
/ SOFTWARE: FastSeq for Windows Version 2.0b  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/385,540A  
/ FILING DATE:  
/ CLASSIFICATION: 424  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: 08/600,273  
/ FILING DATE: 08-FEB-1996  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Bloom, Allen  
/ REGISTRATION NUMBER: 29,135  
/ REFERENCE/DOCKET NUMBER: 314572-101A  
/ TELEPHONE: 609-520-3214  
/ TELEFAX: 609-520-3259  
/ TELEX:  
/ INFORMATION FOR SEQ ID NO: 19:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 25 amino acids  
/ TYPE: amino acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ US-08-385-540A-19

Query Match 94.7%; Score 90; DB 2; Length 25;  
Best Local Similarity 85.0%; Pred. No. 8,7e-10;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYCXKHSHQAHQVSLNS 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 IVGGYECTKHSQAHQVSLNS 20

RESULT 3  
US-08-385-540A-20  
/ Sequence 20, Application US/08385540A  
/ Patent No. 5945102  
/ GENERAL INFORMATION:  
/ APPLICANT: de Faire, Johan  
/ TITLE OF INVENTION: Wound Care With Multifunctional  
/ TITLE OF INVENTION: Enzyme  
/ NUMBER OF SEQUENCES: 20  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Dechert Price & Rhoads  
/ STREET: 997 Lenox Drive, Building 3, Suite 210  
/ CITY: Lawrenceville  
/ STATE: NJ  
/ COUNTRY: USA  
/ ZIP: 08543  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Diskette  
/ COMPUTER: IBM Compatible  
/ OPERATING SYSTEM: Windows 95  
/ SOFTWARE: FastSeq for Windows Version 2.0b  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/385,540A  
/ FILING DATE:  
/ CLASSIFICATION: 424  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: 08/600,273  
/ FILING DATE: 08-FEB-1996  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Bloom, Allen

/ REGISTRATION NUMBER: 29,135  
/ REFERENCE/DOCKET NUMBER: 314572-101A  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: 609-520-3214  
/ TELEFAX: 609-520-3259  
/ TELEX:  
/ INFORMATION FOR SEQ ID NO: 20:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 25 amino acids  
/ TYPE: amino acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ US-08-385-540A-20

Query Match 94.7%; Score 90; DB 2; Length 25;  
Best Local Similarity 85.0%; Pred. No. 8,7e-10;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYCXKHSHQAHQVSLNS 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 IVGGYECTKHSQAHQVSLNS 20

RESULT 4  
US-08-600-273A-19  
/ Sequence 19, Application US/08600273A  
/ Patent No. 5958406  
/ GENERAL INFORMATION:  
/ APPLICANT: de Faire, Johan  
/ APPLICANT: Franklin, Richard L.  
/ APPLICANT: Kay, John  
/ TITLE OF INVENTION: Acne Treatment With Multifunctional  
/ TITLE OF INVENTION: Enzyme  
/ NUMBER OF SEQUENCES: 20  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Dechert Price & Rhoads  
/ STREET: 997 Lenox Drive, Building 3, Suite 210  
/ CITY: Lawrenceville  
/ STATE: NJ  
/ COUNTRY: USA  
/ ZIP: 08543  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Diskette  
/ COMPUTER: IBM Compatible  
/ OPERATING SYSTEM: Windows 95  
/ SOFTWARE: FastSeq for Windows Version 2.0b  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/600,273A  
/ FILING DATE: 08-FEB-1996  
/ CLASSIFICATION: 424  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: 08/486,820  
/ FILING DATE: 07-JUN-1995  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: 08/385,540  
/ FILING DATE: 08-FEB-1995  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Bloom, Allen  
/ REGISTRATION NUMBER: 29,135  
/ REFERENCE/DOCKET NUMBER: 314572-101C  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: 609-620-3214  
/ TELEFAX: 609-620-3259  
/ TELEX:  
/ INFORMATION FOR SEQ ID NO: 19:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 25 amino acids  
/ TYPE: amino acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ US-08-600-273A-19

Query Match 94.7%; Score 90; DB 2; Length 25;

Best Local Similarity 85.0%; Pred. No. 8.7e-10;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYXCXXHQAQVSLNS 20  
||||| | ||||| |||||  
DB 1 IVGGYECTKHQAQVSLNS 20

RESULT 5  
US-08-600-273A-20  
; Sequence 20, Application US/08600273A  
; Patent No. 5958406  
; GENERAL INFORMATION:  
; APPLICANT: de Faire, Johan  
; APPLICANT: Franklin, Richard L.  
; TITLE OF INVENTION: Acne Treatment With Multifunctional  
; TITLE OF INVENTION: Enzyme  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert Price & Rhoads  
; STREET: 997 Lenox Drive, Building 3, Suite 210  
; CITY: Lawrenceville  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 08543  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/600,273A  
; FILING DATE: 08-FEB-1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/486,820  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/385,540  
; FILING DATE: 08-FEB-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bloom, Allen  
; REGISTRATION NUMBER: 29,135  
; REFERENCE/DOCKET NUMBER: 314572-101C  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 609-620-3214  
; TELEFAX: 609-620-3259  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 25 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-600-273A-20

Query Match 94.7%; Score 90; DB 2; Length 25;  
Best Local Similarity 85.0%; Pred. No. 8.7e-10;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYXCXXHQAQVSLNS 20  
||||| | ||||| |||||  
DB 1 IVGGYECTKHQAQVSLNS 20

RESULT 6  
US-08-486-820-19  
; Sequence 19, Application US/08486820  
; Patent No. 6030612  
; GENERAL INFORMATION:  
; APPLICANT: de Faire, Johan  
; APPLICANT: Franklin, Richard L.

; APPLICANT: Kay, John  
; TITLE OF INVENTION: Antimicrobial Uses Of Multifunctional  
; TITLE OF INVENTION: Enzyme  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert Price & Rhoads  
; STREET: 997 Lenox Drive, Building 3, Suite 210  
; CITY: Lawrenceville  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 08543  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/486,820  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/486,820  
; FILING DATE: 07-JUNE-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bloom, Allen  
; REGISTRATION NUMBER: 29,135  
; REFERENCE/DOCKET NUMBER: 314572-101B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 609-620-3214  
; TELEFAX: 609-620-3259  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 25 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-486-820-19

Query Match 94.7%; Score 90; DB 3; Length 25;  
Best Local Similarity 85.0%; Pred. No. 8.7e-10;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYXCXXHQAQVSLNS 20  
||||| | ||||| |||||  
DB 1 IVGGYECTKHQAQVSLNS 20

RESULT 7  
US-08-486-820-20  
; Sequence 20, Application US/08486820  
; Patent No. 6030612  
; GENERAL INFORMATION:  
; APPLICANT: de Faire, Johan  
; APPLICANT: Franklin, Richard L.  
; APPLICANT: Kay, John  
; TITLE OF INVENTION: Antimicrobial Uses Of Multifunctional  
; TITLE OF INVENTION: Enzyme  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert Price & Rhoads  
; STREET: 997 Lenox Drive, Building 3, Suite 210  
; CITY: Lawrenceville  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 08543  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/486,820

```
;
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/486,820
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 314572-101B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-620-3214
; TELEFAX: 609-620-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-486-820-20

Query Match 94.7%; Score 90; DB 3; Length 25;
Best Local Similarity 85.0%; Pred. No. 8.7e-10;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 IVGGYCXHXSHQAHQVSLNS 20
| | | | | | | | | | | | | | | | | | | | |
Db 1 IVGGYECTRHSHQAHQVSLNS 20

RESULT 8
US-08-385-540A-18
; Sequence 18, Application US/08385540A
; Patent No. 5945102
; GENERAL INFORMATION:
; APPLICANT: de Faire, Johan
; TITLE OF INVENTION: Wound Care With Multifunctional
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/385,540A
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/486,820
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/385,540
; FILING DATE: 08-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 314572-101C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-620-3214
; TELEFAX: 609-620-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-385-540A-18

Query Match 94.7%; Score 90; DB 3; Length 25;
Best Local Similarity 85.0%; Pred. No. 8.7e-10;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 IVGGYCXHXSHQAHQVSLNS 20
| | | | | | | | | | | | | | | | | | | | |
Db 1 IVGGYECTRHSHQAHQVSLNS 20

RESULT 8
US-08-385-540A-18
; Sequence 18, Application US/08385540A
; Patent No. 5945102
; GENERAL INFORMATION:
; APPLICANT: de Faire, Johan
; TITLE OF INVENTION: Wound Care With Multifunctional
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/385,540A
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/486,820
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/385,540
; FILING DATE: 08-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 314572-101A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-385-540A-18
```

```
Query Match 82.1%; Score 78; DB 2; Length 25;
Best Local Similarity 75.0%; Pred. No. 1.2e-07;
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 IVGGYCXHXSHQAHQVSLNS 20
| | | | | | | | | | | | | | | | | | | | |
Db 1 IVGGYECKAYSQAYQVSLNS 20

RESULT 9
US-08-600-273A-18
; Sequence 18, Application US/08600273A
; Patent No. 5958406
; GENERAL INFORMATION:
; APPLICANT: de Faire, Johan
; APPLICANT: Franklin, Richard L.
; APPLICANT: Kay, John
; TITLE OF INVENTION: Acne Treatment With Multifunctional
; TITLE OF INVENTION: Enzyme
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/600,273A
; FILING DATE: 08-FEB-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/486,820
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/385,540
; FILING DATE: 08-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 314572-101C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-620-3214
; TELEFAX: 609-620-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-600-273A-18

Query Match 82.1%; Score 78; DB 2; Length 25;
Best Local Similarity 75.0%; Pred. No. 1.2e-07;
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 IVGGYCXHXSHQAHQVSLNS 20
| | | | | | | | | | | | | | | | | | | | |
Db 1 IVGGYECKAYSQAYQVSLNS 20

RESULT 10
US-08-486-820-18
; Sequence 18, Application US/08486820
; Patent No. 6030612
; GENERAL INFORMATION:
```

```
; APPLICANT: de Faire, Johan
; APPLICANT: Franklin, Richard L.
; APPLICANT: Kay, John
; TITLE OF INVENTION: Antimicrobial Uses Of Multifunctional
; TITLE OF INVENTION: Enzyme
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,820
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/486,820
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 314572-101B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-620-3214
; TELEFAX: 609-620-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-486-820-18

Query Match      82.1%; Score 78; DB 3; Length 25;
Best Local Similarity 75.0%; Pred. No. 1.2e-07;
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 IVGGYCXHXSHQAHQVSLNS 20
   ||||| | :|||:|||||
Db 1 IVGGYCKAYSQAYQVSLNS 20

RESULT 11
US-09-220-731-17
; Sequence 17, Application US/09220731A
; Patent No. 6232088
; GENERAL INFORMATION:
; APPLICANT: Phairen Medical, Inc.
; APPLICANT: Richard L. Franklin
; APPLICANT: Yves St. Pierre
; TITLE OF INVENTION: Treatment and Prevention of Immune
; TITLE OF INVENTION: Rejection Reactions
; FILE REFERENCE: 314572-101D
; CURRENT APPLICATION NUMBER: US/09/220,731A
; CURRENT FILING DATE: 1998-12-24
; EARLIER APPLICATION NUMBER: PCT/SE93/00455
; EARLIER FILING DATE: 1993-05-21
; EARLIER APPLICATION NUMBER: 08/338,501
; EARLIER FILING DATE: 1994-11-22
; EARLIER APPLICATION NUMBER: 08/385,540
; EARLIER FILING DATE: 1995-02-08
; EARLIER APPLICATION NUMBER: 08/486,820
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/600,273
; EARLIER FILING DATE: 1996-02-08
```

```
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Salmon
US-09-220-731-17

Query Match      82.1%; Score 78; DB 4; Length 25;
Best Local Similarity 75.0%; Pred. No. 1.2e-07;
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 IVGGYCXHXSHQAHQVSLNS 20
   ||||| | :|||:|||||
Db 1 IVGGYCKAYSQAYQVSLNS 20

RESULT 12
US-08-978-404B-44
; Sequence 44, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; TITLE OF INVENTION: FIBRINOGEN
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5968782e
US-08-978-404B-44

Query Match      74.7%; Score 71; DB 2; Length 246;
Best Local Similarity 70.0%; Pred. No. 2.9e-05;
Matches 14; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 IVGGYCXHXSHQAHQVSLNS 20
   ||||| | :|||:|||||
Db 24 IVGGYTCPEHSVPYQVSLNS 43

RESULT 13
US-08-956-267A-2
```

; Sequence 2, Application US/08956267A  
; Patent No. 5945328  
; GENERAL INFORMATION:  
; APPLICANT: WOLDIKE, Helle Fabricius  
; APPLICANT: KJELDSEN, Thomas Borglum  
; TITLE OF INVENTION: A Process For Producing Trypsin  
; TITLE OF INVENTION: (Trypsinogen)  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5945328o No. 5945328disk of No. 5945328th America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/956,267A  
; FILING DATE: 22-OCT-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rozek, Carol  
; REGISTRATION NUMBER: 36,993  
; REFERENCE/DOCKET NUMBER: 4500.204-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 247 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-956-267A-2

Query Match 69.5%; Score 66; DB 2; Length 247;  
Best Local Similarity 65.0%; Pred. No. 0.00023;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 IVGGYXCXHSQAHQVSLNS 20  
||||| : : : : :  
Db 25 IVGGYTCAANSIPYQVSLNS 44

RESULT 14  
US-08-467-155A-10  
; Sequence 10, Application US/08467155A  
; Patent No. 5736377  
; GENERAL INFORMATION:  
; APPLICANT: Band, Vimla  
; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
; TITLE OF INVENTION: MOLECULES AND METHODS  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467,155A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00398/100001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 271 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-467-155A-10

Query Match 69.5%; Score 66; DB 1; Length 271;  
Best Local Similarity 53.6%; Pred. No. 0.00026;  
Matches 15; Conservative 1; Mismatches 4; Indels 8; Gaps 1;

Qy 1 IVGGYX-----CXHSQAHQVSLNS 20  
||||| : : : : :  
Db 10 IVGGYSSTRYPHCKAYSQPHQVSLNS 37

RESULT 15  
US-08-628-198-10  
; Sequence 10, Application US/08628198  
; Patent No. 5843694  
; GENERAL INFORMATION:  
; APPLICANT: Band, Vimla  
; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
; TITLE OF INVENTION: MOLECULES AND METHODS  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/628,198  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/467,155  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00398/100002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 271 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-628-198-10

Query Match 69.5%; Score 66; DB 2; Length 271;  
Best Local Similarity 53.6%; Pred. No. 0.00026;



Matches 15; Conservative 1; Mismatches 4; Indels 8; Gaps 1;

Qy 1 IVGGYX-----CXXHSQAHQVSLNS 20  
||||| | : || || || || ||  
Db 10 IVGGYSSTRYPPIECKAYSQPHQVSLNS 37

Search completed: February 12, 2003, 10:30:06  
Job time : 5.29851 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:16:06 ; Search time 14.1493 Seconds  
(without alignments)  
291.248 Million cell updates/sec

Title: US-10-036-371-1

Perfect score: 95

Sequence: 1 IVGGYCXHXSHQHQVSLNS 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*

1: sp.archaea.\*

2: sp.bacteria.\*

3: sp.fungi.\*

4: sp.human.\*

5: sp.invertebrate.\*

6: sp.mammal.\*

7: sp.mhc.\*

8: sp.organella.\*

9: sp.phage.\*

10: sp.plant.\*

11: sp.rodent.\*

12: sp.virus.\*

13: sp.unclassified.\*

14: sp.vertibrate.\*

15: sp.rvirus.\*

16: sp.bacteriap.\*

17: sp.archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	90	94.7	219	13	Q91036 gadus morhua
2	89	93.7	117	13	Q9DFJ6
3	89	93.7	242	13	Q93266
4	79	83.2	240	13	Q98TH0
5	78	82.1	238	13	Q9W7Q6
6	78	82.1	242	13	Q9W7Q7
7	77	81.1	241	13	Q98TG9
8	72	75.8	244	13	Q42159
9	72	75.8	245	13	Q42160
10	71	74.7	247	13	Q9W7Q5
11	69	72.6	237	13	Q91515
12	69	72.6	242	13	Q92099
13	67	70.5	247	11	Q9CPN7
14	66	69.5	244	13	Q8QGM3
15	65	68.4	246	11	Q9Z1R9
16	65	68.4	246	11	Q9R0T7

17	64	67.4	246	11	Q9QUK9
18	63	66.3	247	13	Q42158
19	63	66.3	247	13	Q42608
20	62	65.3	247	11	Q9CPN9
21	61	64.2	249	13	Q9W6K0
22	60	63.2	249	13	Q92046
23	60	63.2	673	13	Q9W6J8
24	58	61.1	247	11	Q9D7Y7
25	57	60.0	250	13	Q93265
26	57	60.0	344	13	Q9W6J9
27	52	54.7	48	4	Q9UE84
28	52	54.7	176	4	Q15098
29	52	54.7	180	4	Q96A30
30	52	54.7	195	4	Q07277
31	52	54.7	234	4	Q15096
32	52	54.7	261	6	Q9NIQ1
33	52	54.7	262	4	Q8TCV8
34	51	53.7	235	11	Q63274
35	51	53.7	249	11	Q9QYN4
36	51	53.7	276	11	Q9QYN3
37	50	52.6	17	11	Q9QVC2
38	50	52.6	26	11	Q9QV96
39	50	52.6	28	11	Q9QV97
40	48	50.5	24	11	Q9QV67
41	46	48.4	239	11	Q63275
42	46	48.4	261	11	Q88309
43	45	47.4	261	6	Q29474
44	45	47.4	265	5	O18488
45	45	47.4	270	5	Q8WR11

## ALIGNMENTS

## RESULT 1

Q91036 PRELIMINARY; PRT; 219 AA.  
ID AC Q91036;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Trypsinogen I (Fragment).  
OS Gadus morhua (Atlantic cod).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.  
OX NCBI\_TaxID=8049;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ong T.L., Armstrong R.F., McNamara P., Buckley L.J.;  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.  
CC EMBL; U47819; AAB02196.1; -.  
DR HSSP; P00763; 1DPO.  
DR MEROPS; S01.125; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser. protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease.  
FT NON\_TER 1  
SQ SEQUENCE 219 AA; 23525 MW; C96964EB49CED1DA CRC64;

Query Match 94.7%; Score 90; DB 13; Length 219;  
Best Local Similarity 85.0%; Pred. No. 2.8e-09;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 IVGGYCXHXSHQHQVSLNS 20

• • •

```

Q9W706
ID Q9W706 PRELIMINARY; PRT; 238 AA.
AC Q9W706;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsinogen 2 (Fragment).
OS Paralicthys olivaceus (Flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidae; Paralicthyidae; Paralicthys.
OX NCBI_TaxID=8255;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RA Suzuki T., Srivastava A.S., Kurokawa T.;
RL "Japanese flounder mRNA for trypsinogen 2.";
RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB029751; BAA82362.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.125; -.
DR InterPro; IPR001314; Chymotrypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT NON TER 1
SQ SEQUENCE 238 AA; 26071 MW; F2B8908058B8D062 CRC64;

Query Match 82.1%; Score 78; DB 13; Length 238;
Best Local Similarity 75.0%; Pred. No. 6.3e-07;
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 IVGGYXCXHXSOAHQVSLNS 20
DB 18 IVGGYECTPYSPHQVSLNS 37
||||| | : |||||

RESULT 6
Q9W707
ID Q9W707 PRELIMINARY; PRT; 242 AA.
AC Q9W707;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsinogen 1.
OS Paralicthys olivaceus (Flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidae; Paralicthyidae; Paralicthys.
OX NCBI_TaxID=8255;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RA Suzuki T., Srivastava A.S., Kurokawa T.;
RL "Japanese flounder mRNA for trypsinogen 1.";
RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB029750; BAA82362.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.125; -.
DR InterPro; IPR001314; Chymotrypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT NON TER 1
SQ SEQUENCE 242 AA; 26548 MW; 6DA722C80BC194A2 CRC64;

Query Match 82.1%; Score 78; DB 13; Length 242;
Best Local Similarity 75.0%; Pred. No. 6.4e-07;
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 IVGGYXCXHXSOAHQVSLNS 20
DB 21 IVGGYECTPYSPHQVSLNS 40
||||| | : |||||

RESULT 7
Q98TG9
ID Q98TG9 PRELIMINARY; PRT; 241 AA.
AC Q98TG9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsinogen II.
OS Engraulis japonicus (Japanese anchovy).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Clupeomorpha; Engraulidae;
OC Engraulis.
OX NCBI_TaxID=42892;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PYLORIC CAECA;
RA Watabe S., Ansan M.N., Funabara D.;
RL "Anchovy trypsinogen mRNA.";
RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB041930; BAB40330.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.258; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 241 AA; 26282 MW; FE362D39CAEBB2F6 CRC64;

Query Match 81.1%; Score 77; DB 13; Length 241;
Best Local Similarity 75.0%; Pred. No. 1e-06;
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 IVGGYXCXHXSOAHQVSLNS 20
DB 20 IVGGYECTPYSPHQVSLNS 39
||||| | : |||||

RESULT 8
O42159
ID O42159 PRELIMINARY; PRT; 244 AA.
AC O42159;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsinogen B1 precursor (Fragment).
GN TRYPB1.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;

```

```

OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN [1]
RP SEQUENCE FROM N.A.
RA Roach J.C.;
RT "The Molecular Evolution of the Vertebrate Trypsinogens.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AF011900; AAB69656.1; -.
DR HSSP; P00763; LDPO.
DR MEROPS; S01.128; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Signal.
FT NON_TER 1
FT SIGNAL <1 12
FT CHAIN 13 244
FT SIGNAL 244
SQ SEQUENCE 244 AA; 25903 MW; C4582EE07E3B8007 CRC64;

Query Match 75.8%; Score 72; DB 13; Length 244;
Best Local Similarity 73.7%; Pred. No. 9.4e-06;
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVGGYCXHXHQAQVSLN 19
DB 21 IVGGYCAHSPQWQVSLN 39
||||| |||||

RESULT 9
O42160 PRELIMINARY; PRT; 245 AA.
AC O42160;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsinogen b2 precursor (Fragment).
GN TRYPB2.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN [1]
RP SEQUENCE FROM N.A.
RA Roach J.C.;
RT "The Molecular Evolution of the Vertebrate Trypsinogens.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AF011901; AAB69657.1; -.
DR HSSP; P00763; LDPO.
DR MEROPS; S01.128; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Signal.
FT NON_TER 1
FT SIGNAL <1 13
FT CHAIN 14 245
FT SIGNAL 245
SQ SEQUENCE 245 AA; 26001 MW; 9A932508896C93E CRC64;

Query Match 75.8%; Score 72; DB 13; Length 245;

```

```

Best Local Similarity 73.7%; Pred. No. 9.4e-06;
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVGGYCXHXHQAQVSLN 19
DB 22 IVGGYCAHSPQWQVSLN 40
||||| |||||

RESULT 10
Q9W7Q5 PRELIMINARY; PRT; 247 AA.
AC Q9W7Q5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsinogen 3.
OS Paralichthys olivaceus (Flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidae; Paralichthyidae; Paralichthys.
OX NCBI_TaxID=8255;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RA Suzuki T., Srivastava A.S., Kurokawa T.;
RT "Japanese flounder mRNA for trypsinogen 3.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB029752; BAA82364.2; -.
DR HSSP; P00763; LDPO.
DR MEROPS; S01.124; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 247 AA; 26948 MW; DC4B647179DD972 CRC64;

Query Match 74.7%; Score 71; DB 13; Length 247;
Best Local Similarity 73.7%; Pred. No. 1.5e-05;
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVGGYCXHXHQAQVSLN 19
DB 22 IVGGYCAHSPQWQVSLN 40
||||| |||||

RESULT 11
Q91515 PRELIMINARY; PRT; 237 AA.
AC Q91515;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsinogen (Fragment).
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang K., Gan L., Lee I., Roach J., Hood L.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.

```

```

DR EMBL; U25747; AAA75001.1; -.
DR HSSP; P15031; 1BIT.
DR MEROPS; S01.125; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease.
DR NON_TER 1
FT SIGNAL 1 13 POTENTIAL.
SQ SEQUENCE 237 AA; 25726 MW; 30D2DBAAC39080C2 CRC64;

Query Match 72.6%; Score 69; DB 13; Length 242;
Best Local Similarity 70.0%; Pred. No. 3.5e-05;
Matches 14; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVGGYKXCHXSHQAHQVSLNS 20
DB 16 IVGGTECKNSVAYQVSLNS 35

RESULT 12
Q92099 ID Q92099 PRELIMINARY; PRT; 242 AA.
AC Q92099;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Trypsin precursor (EC 3.4.21.4).
OS Paratubothenia magellanica (Maoi cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Notothenioidei; Nototheniidae; Paratubothenia.
OX NCBI_TaxID=37005;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/C;
RX MEDLINE=21103195; PubMed=8949488;
RA Chen F., Rowen L., Hood L., Rothenberg E.V.;
RT "Differential transcriptional regulation of individual TCR Vbeta segments before gene rearrangement.";
RL J. Immunol. 166:1771-1780(2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.
DR EMBL; AK007406; BAB25018.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.129; -.
DR MGD; MGI:1920876; 181009J06Rik.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease; signal.
FT SIGNAL 1 13 POTENTIAL.
SQ SEQUENCE 242 AA; 26201 MW; 3F4DE7CE80C4477C CRC64;

Query Match 72.6%; Score 69; DB 13; Length 242;
Best Local Similarity 70.0%; Pred. No. 3.5e-05;
Matches 14; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVGGYKXCHXSHQAHQVSLNS 20
DB 21 IVGGTECKNSVAYQVSLNS 40

RESULT 13
Q92099 ID Q92099 PRELIMINARY; PRT; 247 AA.
AC Q92099;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 181009J06Rik protein (Trypsinogen 4).
GN 181009J06Rik OR TRYPSINOGEN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszewski A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/C;
RX MEDLINE=21103195; PubMed=11160223;
RA Chen F., Rowen L., Hood L., Rothenberg E.V.;
RT "Differential transcriptional regulation of individual TCR Vbeta segments before gene rearrangement.";
RL J. Immunol. 166:1771-1780(2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.
DR EMBL; AK007406; BAB25018.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.129; -.
DR MGD; MGI:1920876; 181009J06Rik.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease.
FT SIGNAL 1 13 POTENTIAL.
SQ SEQUENCE 247 AA; 26503 MW; EDF5F7696833C7BC CRC64;

Query Match 70.5%; Score 67; DB 11; Length 247;
Best Local Similarity 68.4%; Pred. No. 8.8e-05;
Matches 13; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVGGYKXCHXSHQAHQVSLN 19
DB 24 IVGGYTCPKHVPYQVSLN 42

RESULT 14
Q8QGW3

```

```
ID Q8QGW3 PRELIMINARY; PRT; 244 AA.
AC Q8QGW3;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Trypsinogen (EC 3.4.21.4).
GN TRY.
OS Anguilla japonica (Japanese eel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguillidae;
OC Anguilla.
OX NCBI_TaxID=7937;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RA Kurokawa T., Suzuki T., Ohta H., Kagawa H., Tanaka H., Unuma T.;
RT "Expression of pancreatic enzyme genes during the early larval stage
of Japanese eel, Anguilla japonica.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB070720; BAB85634.1; -.
KW Hydrolase.
SQ SEQUENCE. 244 AA; 26317 MW; 0EB3B68E8706D52D CRC64;

Query Match 69.5%; Score 66; DB 13; Length 244;
Best Local Similarity 65.0%; Pred. No. 0.00014;
Matches 13; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 IVGGYXCXHXSOAHQVSLNS 20
DB 21 IVGGYCEPHSQPWQASLNA 40

RESULT 15
Q9Z1R9 PRELIMINARY; PRT; 246 AA.
AC Q9Z1R9;
DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Trypsinogen 16
GN TRYGN16 OR TRYPSINOGEN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA Rowen L., Hood L.;
RT "Comparison between strains Balb/C and 129 in a region of the mouse T
cell receptor beta locus.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=21103195; PubMed=11160223;
RA Chen F., Rowen L., Hood L., Rothenberg E.V.;
RT "Differential transcriptional regulation of individual TCR Vbeta
segments before gene rearrangement.";
RL J. Immunol. 166:1771-1780(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
CC EMBL; AF107342; AAC79093.1; -.
DR EMBL; AE000665; AAB69088.1; -.
DR HSSP; P00763; LDPO.
DR MEROPS; S01.063; -.
DR MGD; MGI:2148749; Trygn16.
DR InterPro; IPR001314; Chymotrypsin.
DR Pfam; PF00089; trypsin; 1
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SMO0020; Tryp_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
```

```
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 246 AA; 26134 MW; 34E173B18CA2F463 CRC64;

Query Match 68.4%; Score 65; DB 11; Length 246;
Best Local Similarity 65.0%; Pred. No. 0.00021;
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVGGYXCXHXSOAHQVSLNS 20
DB 24 IVGGYTCRENSVPYQVSLNS 43

Search completed: February 12, 2003, 10:27:17
Job time : 16.1493 secs
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:04:45 ; Search time 2.68657 Seconds  
(without alignments)  
308.768 Million cell updates/sec

Title: US-10-036-371-1

Perfect score: 95

Sequence: 1 IVGGYCXHXSHQHQVLSNS 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	90	94.7	241	1	TRY1_GADMO
2	90	94.7	241	1	TRYX_GADMO
3	84.2	242	1	1	TRY1_SALSA
4	79	83.2	231	1	TRY2_SALSA
5	71	74.7	246	1	TRY1_RAT
6	71	74.7	248	1	TRY3_CHICK
7	69	72.6	246	1	TRYA_RAT
8	69	72.6	246	1	TRYB_RAT
9	66	69.5	231	1	TRYP_PIG
10	66	69.5	246	1	TRY1_CANFA
11	65	68.4	246	1	TRY2_RAT
12	65	68.4	247	1	TRY3_HUMAN
13	65	68.4	304	1	TRY4_HUMAN
14	64	67.4	247	1	TRY1_HUMAN
15	64	67.4	247	1	TRY2_HUMAN
16	63	66.3	243	1	TRY1_BOVIN
17	63	66.3	247	1	TRY2_BOVIN
18	62	65.3	247	1	TRY2_CANFA
19	62	65.3	247	1	TRY3_RAT
20	62	65.3	248	1	TRY1_CHICK
21	62	65.3	248	1	TRY2_CHICK
22	60	63.2	246	1	TRY2_MOUSE
23	57	60.0	238	1	TRY3_SALSA
24	56	58.9	244	1	TRY2_XENLA
25	56	58.9	247	1	TRY4_RAT
26	55	57.9	261	1	TRY2_HUMAN
27	54	56.8	229	1	TRYP_SQUAC
28	54	56.8	250	1	TRYP_PLEPL
29	54	56.8	261	1	TRY1_RAT
30	52	54.7	244	1	TRY2_HUMAN
31	52	54.7	261	1	TRY3_HUMAN
32	52	54.7	261	1	TRY3_MACMU
33	52	54.7	262	1	TRY1_HUMAN

RESULT 1					
TRY1_GADMO					
ID	TRY1_GADMO	STANDARD;	PRT;	241	AA.
AC	P16049;	Q91040;	Q92156;		
DT	01-APR-1990	(Rel. 14, Created)			
DT	01-NOV-1997	(Rel. 35, Last sequence update)			
DT	15-JUN-2002	(Rel. 41, Last annotation update)			
DE	Trypsin I precursor (EC 3.4.21.4).				
OS	Gadus morhua (Atlantic cod).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;				
OC	Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.				
OX	NCBI_TaxID=8049;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Pyeloric caeca;				
RX	MEDLINE=94039130; PubMed=8223632;				
RA	Gudmundsdottir A., Gudmundsdottir E., Oskarsson S., Bjarnason J.B.,				
RA	Eakin A.E., Craik C.S.;				
RT	"Isolation and characterization of cDNAs from Atlantic cod encoding				
RT	two different forms of trypsinogen."				
RL	Eur. J. Biochem. 217:1091-1097(1993).				
RN	[2]				
RP	SEQUENCE OF 20-58.				
RC	TISSUE=Pyeloric caeca;				
RX	MEDLINE=89210867; PubMed=2707266;				
RA	Asgeirsson B., Fox J.W., Bjarnason J.B.;				
RT	"Purification and characterization of trypsin from the poikilotherm				
RT	Gadus morhua".				
RL	Eur. J. Biochem. 180:85-94(1989).				
CC	-!- CATALYTIC ACTIVITY: Preferential cleavage: Arg- -Xaa, Lys- -Xaa.				
CC	-!- SUBCELLULAR LOCATION: Extracellular.				
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.				
-----					
This SWISS-PROT entry is copyright. It is produced through a collaboration					
between the Swiss Institute of Bioinformatics and the EMBL Outstation -					
the European Bioinformatics Institute. There are no restrictions on its					
use by non-profit institutions as long as its content is in no way					
modified and this statement is not removed. Usage by and for commercial					
entities requires a license agreement (see http://www.isb-sib.ch/announce/					
or send an email to license@isb-sib.ch).					
-----					
EMBL;	X76886;	CAA54214.1;	-		
PIR;	S03570;	S03570.			
HSSP;	P00763;	1DPO.			
DR	MEROPS;	S01.151;	-		
DR	InterPro;	IPR001314;	Chymotrypsin.		
DR	InterPro;	IPR001254;	Ser. protease_Try.		
DR	Pfam;	PF00089;	trypsin;	1.	
DR	PRINTS;	PR00722;	CHYMOTRYPSIN.		
DR	SMART;	SM00020;	Tryp_SPC;	1.	
DR	PROSITE;	P850240;	TRYPSIN_DOM;	1.	
DR	PROSITE;	P300134;	TRYPSIN_HIS;	1.	
DR	PROSITE;	P300135;	TRYPSIN_SER;	1.	
KW	Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal; Multigene family.				

```

FT SIGNAL 1 13 POTENTIAL.
FT PROPEP 14 19 ACTIVATION PEPTIDE.
FT CHAIN 20 241 TRYPSIN I.
FT ACT_SITE 59 59 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 103 103 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 195 195 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 26 155 BY SIMILARITY.
FT DISULFID 44 60 BY SIMILARITY.
FT DISULFID 128 228 BY SIMILARITY.
FT DISULFID 135 201 BY SIMILARITY.
FT DISULFID 166 180 BY SIMILARITY.
FT DISULFID 191 215 BY SIMILARITY.
FT SITE 189 189 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
FT SITE 25 25 E -> Q (IN REF. 2).
FT CONFLICT 27 28 TK -> EA (IN REF. 2).
FT CONFLICT 43 43 F -> Y (IN REF. 2).
FT CONFLICT 49 52 VSKD -> IN (IN REF. 2).
SQ SEQUENCE 241 AA; 44EC9A0106AD1A68 CRC64;

Query Match 94.7%; Score 90; DB 1; Length 241;
Best Local Similarity 85.0%; Pred. No. 1.5e-09;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYXCXHSQAHQVSLNS 20
| | | | | | | | | | | | | | | | | | | | |
Db 20 IVGGYECTRHQAHQVSLNS 39

RESULT 2
TRYX_GADMO STANDARD; PRT; 241 AA.
AC Q91041;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin X precursor (EC 3.4.21.4).
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.
OX NCBI_TaxID=8049;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pyloxic caeca;
RX MEDLINE=94039130; PubMed=8223632;
RA Gudmundsdottir A., Gudmundsdottir E., Oskarsen S., Bjarnason J.B.,
RA Eakin A.E., Craik C.S.;
RT "Isolation and characterization of cDNAs from Atlantic cod encoding
RT two different forms of trypsinogen."
RL Eur. J. Biochem. 217:1091-1097(1993).
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X76887; CAA54215.1; -.
CC HSSP; P00763; IDPO.
CC
CC MEROPS; S01.151; -.
CC InterPro; IPR001314; Chymotrypsin.
CC InterPro; IPR001254; Ser_protease_Try.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; Tryp_SPC; 1.
CC PROSITE; PS00240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.

```

```

KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
KW Multigene family.
FT SIGNAL 1 13 POTENTIAL.
FT PROPEP 14 19 ACTIVATION PEPTIDE.
FT CHAIN 20 241 TRYPSIN X.
FT ACT_SITE 59 59 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 103 103 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 195 195 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 26 155 BY SIMILARITY.
FT DISULFID 44 60 BY SIMILARITY.
FT DISULFID 128 228 BY SIMILARITY.
FT DISULFID 135 201 BY SIMILARITY.
FT DISULFID 166 180 BY SIMILARITY.
FT DISULFID 191 215 BY SIMILARITY.
FT SITE 189 189 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 241 AA; 25976 MW; 853D7C26BCAF9DD7 CRC64;

Query Match 94.7%; Score 90; DB 1; Length 241;
Best Local Similarity 85.0%; Pred. No. 1.5e-09;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYXCXHSQAHQVSLNS 20
| | | | | | | | | | | | | | | | | | | | |
Db 20 IVGGYECTRHQAHQVSLNS 39

RESULT 3
TRY1_SALSA STANDARD; PRT; 242 AA.
AC P35031;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin I precursor (EC 3.4.21.4).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=96035908; PubMed=7556223;
RA Male R., Lorens J.B., Smals A.O., Torrisen K.R.;
RT "Molecular cloning and characterization of anionic and cationic
RT variants of trypsin from Atlantic salmon."
RL Eur. J. Biochem. 232:677-685(1995).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.82 ANGSTROMS).
RA Smalas A.O., Hordvik A.;
RT "Structure determination and refinement of benzamide-inhibited
RT trypsin from the North Atlantic salmon (Salmo salar) at 1.82-A
RT resolution."
RL Acta Crystallogr. D 49:318-330(1993).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.82 ANGSTROMS).
RX MEDLINE=95148598; PubMed=7845025;
RA Smalas A.O., Heimstad E.S., Hordvik A., Willassen N.P., Male R.;
RT "Cold adaptation of enzymes: structural comparison between salmon and
RT bovine trypsins."
RL Proteins 20:149-166(1994).
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

ENBL; X70073; CAA49678.1; -;  
 PIR; S31778; S31778.  
 HSP; P35031; 1BIT.  
 MEROPS; S01.258; -;  
 InterPro; IPR001254; Ser\_protease\_Try.  
 Pfam; PF00089; trypsin; 1.  
 SMART; SM0020; Tryp\_SPC; 1.  
 PROSITE; PS02040; TRYPSIN\_DOM; 1.  
 PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 PROSITE; PS00135; TRYPSIN\_SER; 1.  
 Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;  
 Multigene family.  
 NON\_TER 1  
 FT SIGNAL 1  
 FT PROPEP 4  
 FT CHAIN 10 231  
 FT ACT\_SITE 49 49  
 FT ACT\_SITE 93 93  
 FT ACT\_SITE 185 185  
 FT DISULFID 16 145  
 FT DISULFID 34 50  
 FT DISULFID 118 218  
 FT DISULFID 125 191  
 FT DISULFID 156 170  
 FT DISULFID 181 205  
 FT SITE 179 179  
 FT SITE 231 AA; 24823 MW; C54A1CAFET4FAE18 CRC64;  
 SQ SEQUENCE 231 AA; 24823 MW; C54A1CAFET4FAE18 CRC64;  
 Query Match 83.2%; Score 79; DB 1; Length 231;  
 Best Local Similarity 75.0%; Pred. No. 1.7e-07;  
 Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 IVGGYKXXHSQAHQVSLNS 20  
 ||||| :||| |||||  
 Db 10 IVGGYCKAYSQPHQVSLNS 29  
 ||||| :||| |||||

RESULT 5  
 TRY1 RAT  
 ID TRY1 RAT STANDARD; PRT; 246 AA.  
 AC P00762;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Trypsin I, anionic precursor (EC 3.4.21.4) (Pretrypsinogen 1).  
 GN TRY1.  
 OS Eukarya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 RN NCBI\_TaxID=10116;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RX MEDLINE=82265624; PubMed=696710;  
 RA McDonald R.J., Stary S.J., Swift G.H.;  
 RT "Two similar but nonallelic rat pancreatic trypsinogens. Nucleotide  
 sequences of the cloned cDNAs.";  
 RL J. Biol. Chem. 257:9724-9732(1982).  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85054880; PubMed=6094547;  
 RA Craik C.S., Choo Q.L., Swift G.H., Quinto C., McDonald R.J.,  
 RA Rutter W.J.;  
 RT "Structure of two related rat pancreatic trypsin genes.";  
 RL J. Biol. Chem. 259:14255-14264(1984).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
 RX MEDLINE=87292123; PubMed=3112942;  
 RA Sprang A., Standing T., Fletcher R.J., Stroud R.M., Finer-Moore J.,

RA Xiong N.-H., Hamlin R., Rutter W.J., Craik C.S.;  
RT "The three-dimensional structure of Asn102 mutant of trypsin: role of  
RT Asp102 in serine protease catalysis.";  
RL Science 237:905-909(1987).  
CC -|- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
CC -|- SUBCELLULAR LOCATION: Extracellular.  
CC -|- MISCELLANEOUS: THIS SEQUENCE REPRESENTS THE PRECURSOR OF THE MAJOR  
CC FORM OF TRYPSIN PRODUCED BY THE ADULT PANCREAS.  
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; V01273; CAA24580.1; -;  
DR EMBL; J00778; AAA98518.1; -;  
DR PIR; A00948; TRRT1.  
DR PDB; 1TRM; 15-JUL-93.  
DR PDB; 2TRM; 16-JUL-88.  
DR PDB; 1BRA; 30-APR-94.  
DR PDB; 1BRB; 31-JUL-94.  
DR PDB; 1BRC; 31-MAY-94.  
DR MEROPS; S01.094; -;  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
DR Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen; Signal;  
KW Multigene family; 3D-structure.  
FT SIGNAL 1 15  
FT PROPEP 16 23 ACTIVATION PEPTIDE.  
FT CHAIN 24 246 TRYPSIN I, ANIONIC.  
FT ACT\_SITE 63 63 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 107 107 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 200 200 CHARGE RELAY SYSTEM.  
FT DISULFID 30 160  
FT DISULFID 48 64  
FT DISULFID 132 233  
FT DISULFID 139 206  
FT DISULFID 171 185  
FT DISULFID 196 220  
FT SITE 194 194  
FT STRAND 25 25  
FT STRAND 28 29  
FT STRAND 32 33  
FT TURN 36 37  
FT TURN 37 37  
FT STRAND 38 42  
FT STRAND 46 54  
FT TURN 55 56  
FT STRAND 57 60  
FT STRAND 62 64  
FT HELIX 62 64  
FT STRAND 70 73  
FT STRAND 77 77  
FT TURN 78 79  
FT STRAND 86 95  
FT TURN 97 98  
FT STRAND 100 100  
FT TURN 101 104  
FT STRAND 105 105  
FT TURN 106 106  
FT STRAND 109 113  
FT STRAND 127 127  
FT TURN 135 136  
FT TURN 138 143  
FT STRAND 157 157

FT STRAND 159 165  
FT HELIX 168 174  
FT TURN 176 178  
FT TURN 181 182  
FT STRAND 183 186  
FT TURN 189 190  
FT STRAND 194 194  
FT TURN 197 198  
FT TURN 200 201  
FT STRAND 203 206  
FT TURN 207 208  
FT STRAND 209 216  
FT TURN 223 224  
FT STRAND 227 231  
FT HELIX 232 234  
FT HELIX 236 244  
FT TURN 245 245  
SQ SEQUENCE 246 AA; 25959 MW; 6AFA0ADAD11943FB5 CRC64;  
  
Query Match 74.7%; Score 71; DB 1; Length 246;  
Best Local Similarity 70.0%; Pred. No. 5.6e-06;  
Matches 14; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
  
QY 1 IVGGYXCXKHSQAHQVSLNS 20  
| | | | | : | | | | |  
DB 24 IVGGYTCPEHSPVQVSLNS 43  
  
RESULT 6  
TRY3-CHICK STANDARD; PRT; 248 AA.  
AC Q90629;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DE 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Trypsin II-P29 precursor (EC 3.4.21.4).  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RX MEDLINE=95251611; PubMed=7733885;  
RA Wang K., Gan L., Lee I., Hood L.E.;  
RT Isolation and characterization of the chicken trypsinogen gene  
family.";  
RL Biochem. J. 307:471-479(1995).  
CC -|- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
CC -|- SUBCELLULAR LOCATION: Extracellular.  
CC -|- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE PANCREAS WHILE  
CC LOWER LEVELS ARE FOUND IN THE LIVER, SPLEEN AND THYMUS.  
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U15157; AAA79914.1; -;  
DR HSP; P00763; IDPO.  
DR MEROPS; S01.151; -;  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.

DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;  
 FT Multigene family.  
 FT SIGNAL 1 16 BY SIMILARITY.  
 FT PROPEP 17 25 ACTIVATION PEPTIDE (BY SIMILARITY).  
 FT CHAIN 26 248 TRYPSIN 11-P29.  
 FT ACT\_SITE 65 65 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 32 162 BY SIMILARITY.  
 FT DISULFID 50 66 BY SIMILARITY.  
 FT DISULFID 134 235 BY SIMILARITY.  
 FT DISULFID 141 208 BY SIMILARITY.  
 FT DISULFID 173 187 BY SIMILARITY.  
 FT DISULFID 198 222 BY SIMILARITY.  
 FT SITE 196 196 REQUIRED FOR SPECIFICITY (BY SIMILARITY).  
 SQ SEQUENCE 248 AA; 26622 MW; E5E16B07622B588E CRC64;

Query Match 74.7%; Score 71; DB 1; Length 248;  
 Best Local Similarity 70.0%; Pred. No. 5.6e-06; Mismatches 5; Indels 0; Gaps 0;  
 Matches 14; Conservative 1;

QY 1 IVGGYXCXHXSHQAHQVSLNS 20  
 ||||| ||| :|||  
 Db 26 IVGGYTCPEHSVPYQVSLNS 45

RESULT 7  
 ID TRYA\_RAT STANDARD; PRT; 246 AA.  
 AC P32821;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Trypsin V-A precursor (EC 3.4.21.4).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=92165057; PubMed=1537555;  
 RA Kang J., Wiegand U., Mueller-Hill B.;  
 RT "Identification of cDNAs encoding two novel rat pancreatic serine  
 proteases";  
 RL Gene 110:181-187(1992).  
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
 CC -!- SUBCELLULAR LOCATION: Extracellular.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X59012; CAA41751.1; -.  
 CC PIR; JQ1471; JQ1471.  
 CC HSSP; P00763; IDPO.  
 CC MEROPS; S01.092; -.  
 CC InterPro; IPR001254; Ser\_protease\_Try.  
 CC PRINTS; P00722; CHYMOTRYPSIN.  
 CC SMART; SM00020; Tryp\_SPC; 1.  
 CC PROSITE; PS0240; TRYPSIN\_DOM; 1.  
 CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 CC PROSITE; PS00135; TRYPSIN\_SER; FALSE NEG.  
 KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;  
 FT Multigene family.

FT SIGNAL 1 15  
 FT PROPEP 16 24 ACTIVATION PEPTIDE.  
 FT CHAIN 25 246 TRYPSIN V-A.  
 FT ACT\_SITE 64 64 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 108 108 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 31 160 BY SIMILARITY.  
 FT DISULFID 49 65 BY SIMILARITY.  
 FT DISULFID 133 233 BY SIMILARITY.  
 FT DISULFID 140 206 BY SIMILARITY.  
 FT DISULFID 171 185 BY SIMILARITY.  
 FT DISULFID 196 220 BY SIMILARITY.  
 FT SITE 194 194 REQUIRED FOR SPECIFICITY (BY SIMILARITY).  
 SQ SEQUENCE 246 AA; 26900 MW; 1EBE59D88BAB1715 CRC64;

Query Match 72.6%; Score 69; DB 1; Length 246;  
 Best Local Similarity 65.0%; Pred. No. 1.3e-05; Mismatches 2; Indels 0; Gaps 0;  
 Matches 13; Conservative 2;

QY 1 IVGGYXCXHXSHQAHQVSLNS 20  
 ||||| ||| :|||  
 Db 25 IVGGYTCPEHSVPYQVSLNA 44

RESULT 8  
 ID TRYB\_RAT STANDARD; PRT; 246 AA.  
 AC P32822;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Trypsin V-B precursor (EC 3.4.21.4).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=92165057; PubMed=1537555;  
 RA Kang J., Wiegand U., Mueller-Hill B.;  
 RT "Identification of cDNAs encoding two novel rat pancreatic serine  
 proteases";  
 RL Gene 110:181-187(1992).  
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
 CC -!- SUBCELLULAR LOCATION: Extracellular.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X59013; CAA41752.1; -.  
 CC PIR; JQ1472; JQ1472.  
 CC HSSP; P00763; IDPO.  
 CC MEROPS; S01.093; -.  
 CC InterPro; IPR001314; Chymotrypsin.  
 CC InterPro; IPR001254; Ser\_protease\_Try.  
 CC Pfam; PF00089; trypsin; 1.  
 CC PRINTS; P00722; CHYMOTRYPSIN.  
 CC SMART; SM00020; Tryp\_SPC; 1.  
 CC PROSITE; PS0240; TRYPSIN\_DOM; 1.  
 CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 CC PROSITE; PS00135; TRYPSIN\_SER; FALSE NEG.  
 KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;  
 FT Multigene family.  
 FT SIGNAL 1 15  
 FT PROPEP 16 24 ACTIVATION PEPTIDE.  
 FT CHAIN 25 246 TRYPSIN V-B.

```

FT ACT SITE 64 64 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 108 108 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 200 200 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 31 160 BY SIMILARITY.
FT DISULFID 49 65 BY SIMILARITY.
FT DISULFID 133 233 BY SIMILARITY.
FT DISULFID 140 206 BY SIMILARITY.
FT DISULFID 171 185 BY SIMILARITY.
FT DISULFID 196 220 BY SIMILARITY.
FT SITE 194 194 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 246 AA; 26819 MW; 1EB899CA1B8A0025 CRC64;

Query Match 72.6%; Score 69; DB 1; Length 246;
Best Local Similarity 65.0%; Pred. No. 1.3e-05;
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVGGYCXHXSHQAHQVSLNS 20
DB 25 IVGGYTCQSHSVYPQVSLNA 44

RESULT 9
TRYP_PIG STANDARD; PRT; 231 AA.
ID TRYP_PIG
AC P00761;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin precursor (EC 3.4.21.4).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE OF 1-10.
RA Charles M., Rovey M., Guidoni A.A., Desnuelle P.;
RT "On trypsinogen and trypsin of pig.";
RL Biochim. Biophys. Acta 69:115-129(1963).
RN [2]
RP SEQUENCE OF 9-231.
RX MEDLINE=73258692; PubMed=4738933;
RA Hermodson M.A., Ericsson L.H., Neurath H., Walsh K.A.;
RT "Determination of the amino acid sequence of porcine trypsin by
sequenator analysis.";
RL Biochemistry 12:3146-3153(1973).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RX MEDLINE=93187998; PubMed=8445634;
RA Huang Q., Liu S., Tang Y.;
RT "Refined 1.6-A resolution crystal structure of the complex formed
between porcine beta-trypsin and MCTI-A, a trypsin inhibitor of the
squash family. Detailed comparison with bovine beta-trypsin and its
complex.";
RL J. Mol. Biol. 229:1022-1030(1993).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RX MEDLINE=92201369; PubMed=1551419;
RA Huang Q., Liu S., Tang Y., Zeng F., Qian R.;
RT "Amino acid sequencing of a trypsin inhibitor by refined 1.6 A X-ray
crystal structure of its complex with porcine beta-trypsin.";
RL FEBS Lett. 297:143-146(1992).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=95035057; PubMed=7947985;
RA Huang Q., Wang Z., Li Y., Liu S., Tang Y.;
RT "Refined 1.8-A resolution crystal structure of the porcine epsilon-
trypsin.";
RL Biochim. Biophys. Acta 1209:77-82(1994).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF COMPLEX WITH LDTI.
RX MEDLINE=97390427; PubMed=9242660;
RA Stubbs M.T., Morenweiser R., Sturzebecher J., Bauer M., Bode W.,
RA Huber R., Piechottka G.P., Matschner G., Sommerhoff C.P., Fritz H.,

```

```

RA Auerswald E.A.;
RT "The three-dimensional structure of recombinant leech-derived
RT trypsin inhibitor in complex with trypsin. Implications for the
RT structure of human mast cell trypsinase and its inhibition.";
RL J. Biol. Chem. 272:19931-19937(1997).
RN [7]
RX X-RAY CRYSTALLOGRAPHY (2.03 ANGSTROMS) OF COMPLEX WITH LDTI.
RX MEDLINE=98046095; PubMed=9384562;
RA di Marco S., Priestle J.P.;
RT "Structure of the complex of leech-derived trypsinase inhibitor (LDTI)
RT with trypsin and modeling of the LDTI-trypsinase system.";
RL Structure 5:1465-1474(1997).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR PIR; A00947; TRPGTR.
DR PDB; 1MCT; 31-JAN-94.
DR PDB; 1AKS; 12-FEB-97.
DR PDB; 1EPT; 07-FEB-95.
DR PDB; 1TFX; 21-JAN-98.
DR PDB; 1LDT; 20-MAY-98.
DR PDB; 1ANI; 01-JUL-98.
DR PDB; 1AVM; 18-NOV-98.
DR PDB; 1AVX; 18-NOV-98.
DR MEROPS; S01.151; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00440; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen;
KW 3D-structure.
FT PROPEP 1 8 ACTIVATION PEPTIDE.
FT CHAIN 9 231 TRYPSIN.
FT ACT_SITE 48 48 CHARGE RELAY SYSTEM.
FT ACT_SITE 92 92 CHARGE RELAY SYSTEM.
FT ACT_SITE 185 185 CHARGE RELAY SYSTEM.
FT DISULFID 15 145
FT DISULFID 33 49
FT DISULFID 117 218
FT DISULFID 124 191
FT DISULFID 156 170
FT DISULFID 181 205
FT SITE 179 179
FT VARIANT 20 20
FT STRAND 10 10
FT STRAND 13 14
FT TURN 17 18
FT TURN 21 22
FT STRAND 23 27
FT STRAND 31 39
FT TURN 40 41
FT STRAND 42 45
FT HELIX 47 49
FT STRAND 55 58
FT STRAND 62 62
FT TURN 63 64
FT STRAND 71 80
FT TURN 82 83
FT TURN 86 88
FT TURN 90 91
FT STRAND 94 98
FT STRAND 112 112
FT TURN 120 121
FT STRAND 123 128
FT STRAND 142 142
FT STRAND 144 150
FT HELIX 153 159
FT TURN 161 163
FT TURN 166 167

```

```

FT STRAND 168 171
FT TURN 174 175
FT STRAND 179 173
FT TURN 182 183
FT TURN 185 186
FT STRAND 188 191
FT TURN 192 193
FT STRAND 194 202
FT TURN 207 207
FT STRAND 208 209
FT STRAND 210 210
FT STRAND 212 216
FT TURN 217 219
FT HELIX 221 230
SQ SEQUENCE 231 AA; 24409 MW; A0A125CF7C138C2 CRC64;

Query Match 69.5%; Score 66; DB 1; Length 231;
Best Local Similarity 65.0%; Pred. No. 4.5e-05;
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVGGYKXKHSHQAHQVSLNS 20
Db 9 IVGGYCAANSIPYQVSLNS 28

RESULT 10
TRY1 CANFA
ID TRY1 CANFA STANDARD; PRT; 246 AA.
AC P06871;
DT 01-JAN-1998 (Rel. 06, Created)
DT 01-JAN-1998 (Rel. 06, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin, cationic precursor (EC 3.4.21.4).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86284628; PubMed=3841794;
RA Pinsky S.D., LaForge K.S., Scheele G.;
RT "Differential regulation of trypsinogen mRNA translation: full-length
RT mRNA sequences encoding two oppositely charged trypsinogen isoenzymes
RT in the dog pancreas."
RL Mol. Cell. Biol. 5:2669-2676(1985).
RL [1]
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M11590; AAA30900.1; -.
CC PIR; B26273; TRDGC.
CC HSSP; P00761; 1EPT.
CC MEROPS; S01.151; -.
CC InterPro; IPR001314; Chymotrypsin.
CC InterPro; IPR001254; Ser. protease_Try.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; Tryp_SPC; 1.
CC PROSITE; PS0240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
CC Multigene family.
CC SIGNAL 1 15 ACTIVATION PEPTIDE.
CC PROPEP 16 23
FT

```

```

FT CHAIN 24 246
FT ACT_SITE 63 63 TRYPSIN, CATIONIC.
FT ACT_SITE 107 107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 200 200 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 30 160 BY SIMILARITY.
FT DISULFID 48 64 BY SIMILARITY.
FT DISULFID 132 233 BY SIMILARITY.
FT DISULFID 139 206 BY SIMILARITY.
FT DISULFID 171 185 BY SIMILARITY.
FT DISULFID 196 220 BY SIMILARITY.
FT SITE 194 194 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 246 AA; 26170 MW; E9E5AIDE2391BBB CRC64;

Query Match 69.5%; Score 66; DB 1; Length 246;
Best Local Similarity 65.0%; Pred. No. 4.8e-05;
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVGGYKXKHSHQAHQVSLNS 20
Db 24 IVGGYTCRSRNVYQVSLNS 43

RESULT 11
TRY2 RAT
ID TRY2 RAT STANDARD; PRT; 246 AA.
AC P00763;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUN-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin II, anionic precursor (EC 3.4.21.4) (Pretrypsinogen II).
GN TRY2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85054880; PubMed=6094547;
RA Craik C.S., Choo Q.L., Swift G.H., Quinto C., McDonald R.J.,
RA Rutter W.J.;
RT "Structure of two related rat pancreatic trypsin genes."
RL J. Biol. Chem. 259:14255-14264(1984).
RN [2]
RP SEQUENCE OF 9-246 FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE-Pancreas;
RX MEDLINE=82265624; PubMed=6896710;
RA McDonald R.J., Stary S.J., Swift G.H.;
RT "Two similar but nonallelic rat pancreatic trypsinogens. Nucleotide
RT sequences of the cloned cDNAs."
RL J. Biol. Chem. 257:9724-9732(1982).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.59 ANGSTROMS).
RX MEDLINE=91351398; PubMed=1881877;
RA Earnest T., Fauman E., Craik C.S., Stroud R.;
RT "1.59-A structure of trypsin at 120 K: comparison of low temperature
RT and room temperature structures."
RL Proteins 10:171-187(1991).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=96214506; PubMed=8634241;
RA Brinen L.S., Willett W.S., Craik C.S., Fletterick R.J.;
RT "X-ray structures of a designed binding site in trypsin show metal-
RT dependent geometry."
RL Biochemistry 35:5999-6009(1996).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC

```





```

RL Gene 136:167-175(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM C).
RA Fukuoka S.,
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- ALTERNATIVE PRODUCTS: 3 isoforms; A (shown here), B and C; seem
CC to be produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: PANCREAS AND BRAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X72781; CAB58178.1; -
DR EMBL; X71345; CAAS0484.1; -
DR EMBL; D45417; BAA08257.1; -
DR PIR; S33496; S33496.
DR HSP; P07477; 1TRN.
DR MEROPS; S01.174; -
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; Trypsin; 1
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; FALSE NEG.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
KW Multigene family; Alternative splicing.
FT SIGNAL 1 ?
FT PROPEP 81 ?
FT CHAIN 81 304
FT ACT_SITE 120 120
FT ACT_SITE 164 164
FT ACT_SITE 257 257
FT DISULFID 87 217
FT DISULFID 105 121
FT DISULFID 196 263
FT DISULFID 228 242
FT DISULFID 253 277
FT SITE 251 251
FT VARSPIC 1 45
FT VARSPIC 1 70
FT CONFLICT 89 89 MISSING (IN REF. 1; CAA50484).
SQ SEQUENCE 304 AA; 32499 MW; 4C4316C31F1D0FC CRC64;

Query Match 68.4%; Score 65; DB 1; Length 304;
Best Local Similarity 65.0%; Pred. No. 9.3e-05;
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 IVGGYXXHQAQVSLNS 20
||||| : : |||||
Db 81 IVGGYTCENSLPYQVSLNS 100

RESULT 14
TRY1_HUMAN
ID TRY1_HUMAN STANDARD; PRT; 247 AA.
AC P07477; Q52955; Q9HAN4; Q9HAN5; Q9HAN6; Q9HAN7;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin I precursor (EC 3.4.21.4) (Cationic trypsinogen).

```

---

```

GN OS PRS1 OR TRY1 OR TRP1 OR TRYPL.
OC Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86221712; PubMed=3011602;
RA Emi M., Nakamura Y., Ogawa M., Yamamoto T., Nishide T., Mori T.,
RA Matsubara K.;
RT "Cloning, characterization and nucleotide sequences of two cDNAs
RT encoding human pancreatic trypsinogens.";
RL Gene 41:305-310(1986).
RN [2]
RP SEQUENCE OF 16-43.
RX MEDLINE=90091010; PubMed=2598466;
RA Kinland M., Russick C., Marks W.H., Borgstrom A.;
RT "Immunoreactive anionic and cationic trypsin in human serum.";
RL Clin. Chim. Acta 184:31-46(1989).
RN [3]
RP SEQUENCE OF 68-151 FROM N.A., AND VARIANT HPC HIS-122.
RX MEDLINE=96438847; PubMed=8841182;
RA Whitcomb D.C., Gorry M.C., Preston R.A., Furey W., Sossenheimer M.J.,
RA Ulrich C.D., Martin S.P., Gates L.K. Jr., Amann S.T., Toskes P.P.,
RA Liddle R., McGrath K., Uomo G., Post J.C., Ehrlich G.D.;
RT "Hereditary pancreatitis is caused by a mutation in the cationic
RT trypsinogen gene.";
RL Nat. Genet. 14:141-145(1996).
RN [4]
RP SEQUENCE OF 15-67 FROM N.A., AND VARIANT HPC GLY-22.
RX MEDLINE=20389982; PubMed=10930381;
RA Teich N., Ockenga J., Hoffmeister A., Manns M., Mossner J., Keim V.;
RT "Chronic pancreatitis associated with an activation peptide mutation
RT that facilitates trypsin activation.";
RL Gastroenterology 119:461-465(2000).
RN [5]
RP SEQUENCE OF 68-151 FROM N.A., AND VARIANTS HPC P-104; C-116 AND F-139.
RA Teich N., Hauer N., Mossner J., Keim V.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS), AND PHOSPHORYLATION.
RX MEDLINE=96266496; PubMed=8683601;
RA Gaboriaud C., Serre L., Guy-Crotte O., Forest E.,
RA Pontecilla-Camps J.-C.;
RT "Crystal structure of human trypsin 1: unexpected phosphorylation of
RT Tyr151.";
RL J. Mol. Biol. 259:995-1010(1996).
RN [7]
RP VARIANTS HPC ILE-29 AND HIS-122.
RX MEDLINE=97463797; PubMed=9322498;
RA Gorry M.C., Gabaizadeh D., Furey W., Gates L.K. Jr., Preston R.A.,
RA Aston C.E., Zhang Y., Ulrich C., Ehrlich G.D., Whitcomb D.C.;
RT "Mutations in the cationic trypsinogen gene are associated with
RT recurrent acute and chronic pancreatitis.";
RL Gastroenterology 113:1063-1068(1997).
RN [8]
RP VARIANT HPC ILE-29.
RX MEDLINE=98295575; PubMed=9633818;
RA Teich N., Mossner J., Keim V.;
RT "Mutations of the cationic trypsinogen in hereditary pancreatitis.";
RL Hum. Mutat. 12:39-43(1998).
RN [9]
RP VARIANTS HPC VAL-16 AND HIS-122.
RX MEDLINE=99315544; PubMed=10381903;
RA Witt H., Luck W., Becker M.;
RT "A signal peptide cleavage site mutation in the cationic trypsinogen
RT gene is strongly associated with chronic pancreatitis.";
RL Gastroenterology 117:7-10(1999).
RN [10]
RP VARIANT HPC ARG-23.
RX MEDLINE=99219545; PubMed=10204851;
RA Ferec C., Ragueneas O., Salomon R., Roche C., Bernard J.P., Guillot M.,
RA Quere I., Faure C., Mercier B., Audrezet M.P., Guillausseau F.J.,

```

RA Dupont C., Munnich A., Bignon J.D., Le Bodic L.;  
 RT "Mutations in the cationic trypsinogen gene and evidence for genetic  
 heterogeneity in hereditary pancreatitis";  
 RL J. Med. Genet. 36:228-232(1999).  
 RN [11]  
 RP VARIANTS HPC THR-29 AND CYS-122.  
 RX MEDLINE=21648565; PubMed=11788572;  
 RA Pfutzer R., Myers E., Applebaum-Shapiro S., Finch R., Ellis I.,  
 RA Neoptolemos J., Kant J.A., Whitcomb D.C.;  
 RT "Novel cationic trypsinogen (PRSS1) N29T and R122C mutations cause  
 autosomal dominant hereditary pancreatitis";  
 RL Gut 50:271-272(2002).  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- MASS SPECTROMETRY: MW=24348; MW\_ERR=2; METHOD=Electropray;  
 CC RANGE=24-247.  
 CC -1- DISEASE: Defects in PRSS1 are a cause of hereditary pancreatitis,  
 (HPC or HP); also known as chronic pancreatitis (CP). HPC is an  
 autosomal dominant disease characterized by the presence of  
 calculi in pancreatic ducts. It causes severe abdominal pain  
 attacks.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M22612; AAA61231.1; --  
 CC EMBL; U70137; AAC50728.1; --  
 CC EMBL; AF314534; AAC30943.1; --  
 CC EMBL; AF315309; AAC30947.1; --  
 CC EMBL; AF315310; AAC30948.1; --  
 CC EMBL; AF315311; AAC30949.1; --  
 CC PIR; A25852; A25852.  
 CC PDB; 1TRN; 03-JUN-95.  
 CC PDB; 1PXY; 17-JUN-98.  
 CC MEROPS; S01.151; --  
 CC Genew; HGNC:9475; PRSS1.  
 CC MIM; 276000; --  
 CC MIM; 167800; --  
 CC InterPro: IPR001314; Chymotrypsin.  
 CC InterPro: IPR001254; Ser. protease\_Try.  
 CC Pfam; PF00089; ctrypsin; 1.  
 CC PRINTS; PR00722; CHYMOTRYPSIN.  
 CC SMART; SM00020; TRYD\_SPC; 1.  
 CC PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
 CC Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;  
 KM Multigene family; Phosphorylation; 3D-structure; Disease mutation.  
 FT SIGNAL 1 15  
 FT PROPEP 16 23  
 FT CHAIN 24 247  
 FT TRYPSIN I.  
 FT ACT\_SITE 63  
 FT CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 107 107  
 FT ACT\_SITE 200 200  
 FT DISULFID 30 160  
 FT DISULFID 48 64  
 FT DISULFID 139 206  
 FT DISULFID 171 185  
 FT DISULFID 196 220  
 FT MOD\_RES 154  
 FT SITE 194 194  
 FT VARIANT 16 16  
 CC PHOSPHORYLATION.  
 CC REQUIRED FOR SPECIFICITY (BY SIMILARITY).  
 CC A -> V (IN HPC; DISRUPTS SIGNAL SEQUENCE  
 CC CLEAVAGE SITE).  
 CC /FTid=VAR\_011693.  
 CC D -> G (IN HPC; INCREASED RATE OF  
 CC ACTIVATION).  
 CC /FTid=VAR\_011652.  
 CC

FT VARIANT 23 23 K -> R (IN HPC; INCREASED RATE OF  
 FT ACTIVATION).  
 FT /FTid=VAR\_011653.  
 FT N -> I (IN HPC).  
 FT /FTid=VAR\_006720.  
 FT N -> T (IN HPC).  
 FT /FTid=VAR\_012712.  
 FT L -> P (IN HPC).  
 FT /FTid=VAR\_011654.  
 FT R -> C (IN HPC).  
 FT /FTid=VAR\_011655.  
 FT R -> C (IN HPC; SUPPRESSES AN  
 FT AUTOCLEAVAGE SITE).  
 FT /FTid=VAR\_012713.  
 FT R -> H (IN HPC; SUPPRESSES AN  
 FT AUTOCLEAVAGE SITE WHICH IS PROBABLY PART  
 FT OF A FAIL-SAFE MECHANISM BY WHICH  
 FT TRYPSIN, WHICH IS ACTIVATED WITHIN THE  
 FT PANCREAS, MAY BE INACTIVATED; LOSS OF  
 FT THIS CLEAVAGE SITE WOULD PERMIT  
 FT AUTODIGESTION RESULTING IN PANCREATITIS).  
 FT /FTid=VAR\_006721.  
 FT C -> F (IN HPC).  
 FT /FTid=VAR\_011656.  
 FT DD49A487B8062813 CRC64;  
 SQ SEQUENCE 247 AA; 26558 MW; 26558 MW;  
 Query Match 67.4%; Score 64; DB 1; Length 247;  
 Best Local Similarity 65.0%; Pred. No. 0.00011;  
 Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 IVGGYKXKHSHQAHQVSLNS 20  
 ||||| : : : : :  
 DB 24 IVGGYCNCENSVPYQVSLNS 43  
 RESULT 15  
 TRY2\_HUMAN STANDARD; PRT; 247 AA.  
 ID TRY2\_HUMAN  
 AC P07478;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-APR-1988 (Rel. 07, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Trypsin II precursor (EC 3.4.21.4) (Anionic trypsinogen).  
 GN PRSS2 OR TRY2 OR TRY2P.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86221712; PubMed=3011602;  
 RA Eni M., Nakamura Y., Ogawa M., Yamamoto T., Nishide T., Mori T.,  
 RA Matsubara K.;  
 RT "Cloning, characterization and nucleotide sequences of two cDNAs  
 encoding human pancreatic trypsinogens.";  
 RL Gene 41:305-310(1986).  
 RN [2]  
 RP SEQUENCE OF 16-49.  
 RX MEDLINE=90091010; PubMed=2598466;  
 RA Kinland M., Russick C., Marks W.H., Borgstrom A.;  
 RA "Immunoreactive anionic and cationic trypsin in human serum.";  
 RL Clin. Chim. Acta 184:31-46(1989).  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

```

CC -----
DR EMBL; M27602; AAA61232.1; -.
DR PIR; B25852; B25852.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.258; -.
DR Genew; HGNC:9483; PRSS2.
DR MIM; 601564; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SMC0020; TRYP_SPE; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR HydroLase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
KW Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 23 ACTIVATION PEPTIDE.
FT CHAIN 24 247 TRYPSIN II.
FT ACT_SITE 63 63 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 107 107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 200 200 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 30 160 BY SIMILARITY.
FT DISULFID 48 64 BY SIMILARITY.
FT DISULFID 171 185 BY SIMILARITY.
FT DISULFID 196 220 BY SIMILARITY.
FT SITE 194 194 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 247 AA; 26488 MW; 82B0F41EB8E3D5DB CRC64;

Query Match 67.4%; Score 64; DB 1; Length 247;
Best Local Similarity 65.0%; Pred. No. 0.00011;
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVGGYKXKHSHQHQVLSNS 20
Db 24 IVGGYCEENSVPYQVLSNS 43

```

Search completed: February 12, 2003, 10:23:12  
 Job time : 3.68657 secs

5 PAGE BLANK (USPTO)



```
RESULT 3
S39048
trypsin (EC 3.4.21.4) X - Atlantic cod
C;Species: Gadus morhua (Atlantic cod)
C;Date: 20-May-1994 #sequence_revision 03-Aug-1995 #text_change 20-Jun-2000
C;Accession: S39048
R;Gudmundsdottir, A.; Gudmundsdottir, E.; Oskareason, S.; Bjarnason, J.B.; Eakin, A.K.; C
Eur. J. Biochem. 217, 1091-1097, 1993
A;Title: Isolation and characterization of cDNAs from Atlantic cod encoding two different
A;Reference number: S39047; MUID:94039130; PMID:8223632
A;Accession: S39048
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-241 <GUD>
A;Cross-references: EMBL:X76887; NID:9450519; PIDN:CAA54215.1; PID:g1334753
C;Superfamily: trypsin; trypsin homology
C;Keywords: hydrolase; serine proteinase
F;1-15/Domain: signal sequence #status predicted <SIG>
F;20-234/Domain: trypsin homology <TRY>
F;59,103,195/Active site: His, Asp, Ser #status predicted
Query Match 94.7%; Score 90; DB 2; Length 241;
Best Local Similarity 85.0%; Pred. No. 1.3e-08;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYXCXHXSHQAHQVSLNS 20
| | | | | : | | | | |
Db 20 IVGGYECTRHSHQAHQVSLNS 39
| | | | | : | | | | |

RESULT 4
S31776
trypsin (EC 3.4.21.4) IA precursor - Atlantic salmon
C;Species: Salmo salar (Atlantic salmon)
C;Date: 03-Mar-1994 #sequence_revision 03-Aug-1995 #text_change 22-Jun-1999
C;Accession: S66659; S31776
R;Male, R.; Lorens, J.B.; Smalas, A.O.; Torrisen, K.R.
Eur. J. Biochem. 232, 677-685, 1995
A;Title: Molecular cloning and characterization of anionic and cationic variants of try
A;Reference number: S66657; MUID:96035908; PMID:7556223
A;Accession: S66659
A;Molecule type: mRNA
A;Residues: 1-242 <MAL>
A;Cross-references: EMBL:X70071; NID:g64381; PIDN:CAA49676.1; PID:g64382
C;Superfamily: trypsin; trypsin homology
C;Keywords: hydrolase; serine proteinase
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-20/Domain: activation peptide #status predicted <APT>
F;21-242/Product: trypsin IA #status predicted <MAT>
F;21-235/Domain: trypsin homology <TRY>
F;27-156,45-61,129-229,136-202,167-181,192-216/Disulfide bonds: #status predicted
F;60,104,196/Active site: His, Asp, Ser #status predicted
Query Match 88.4%; Score 84; DB 2; Length 242;
Best Local Similarity 80.0%; Pred. No. 1.6e-07;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYXCXHXSHQAHQVSLNS 20
| | | | | : | | | | |
Db 21 IVGGYECTRHSHQAHQVSLNS 40
| | | | | : | | | | |

RESULT 5
S31775
trypsin (EC 3.4.21.4) I precursor - Atlantic salmon
C;Species: Salmo salar (Atlantic salmon)
C;Date: 03-Mar-1994 #sequence_revision 03-Aug-1995 #text_change 15-Oct-1999
C;Accession: S66660; S66661; S31775; S31777
R;Male, R.; Lorens, J.B.; Smalas, A.O.; Torrisen, K.R.
Eur. J. Biochem. 232, 677-685, 1995
A;Title: Molecular cloning and characterization of anionic and cationic variants of try
A;Reference number: S66657; MUID:96035908; PMID:7556223
A;Accession: S66660
```

```
A;Molecule type: mRNA
A;Residues: 1-242 <MAL>
A;Cross-references: EMBL:X70075; NID:g64379; PIDN:CAA49680.1; PID:g64380
A;Experimental source: pancreas
A;Accession: S66661
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 28-32,'A',34-242 <MA2>
A;Cross-references: EMBL:X70072; NID:g64383; PIDN:CAA49677.1; PID:g64384
A;Experimental source: pancreas
A;Note: trypsin IB, probably an allelic variant
C;Superfamily: trypsin; trypsin homology
C;Keywords: hydrolase; serine proteinase
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-20/Domain: activation peptide #status predicted <APT>
F;21-242/Product: trypsin I #status predicted <MAT>
F;21-235/Domain: trypsin homology <TRY>
F;27-156,45-61,129-229,136-202,167-181,192-216/Disulfide bonds: #status predicted
F;60,104,196/Active site: His, Asp, Ser #status predicted
Query Match 84.2%; Score 80; DB 2; Length 242;
Best Local Similarity 75.0%; Pred. No. 8.3e-07;
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 IVGGYXCXHXSHQAHQVSLNS 20
| | | | | : | | | | |
Db 21 IVGGYECTKAYSQTHQVSLNS 40
| | | | | : | | | | |

RESULT 6
S31778
trypsin (EC 3.4.21.4) II precursor - Atlantic salmon (fragment)
C;Species: Salmo salar (Atlantic salmon)
C;Date: 22-Nov-1993 #sequence_revision 03-Aug-1995 #text_change 22-Jun-1999
C;Accession: S66658; S31778
R;Male, R.; Lorens, J.B.; Smalas, A.O.; Torrisen, K.R.
Eur. J. Biochem. 232, 677-685, 1995
A;Title: Molecular cloning and characterization of anionic and cationic variants of try
A;Reference number: S66657; MUID:96035908; PMID:7556223
A;Accession: S66658
A;Molecule type: mRNA
A;Residues: 1-231 <MAL>
A;Cross-references: EMBL:X70073; NID:g64385; PIDN:CAA49678.1; PID:g64386
C;Superfamily: trypsin; trypsin homology
C;Keywords: hydrolase; serine proteinase
F;1-4/Domain: signal sequence (fragment) #status predicted <SIG>
F;5-9/Domain: activation peptide #status predicted <APT>
F;10-231/Product: trypsin II #status predicted <MAT>
F;10-224/Domain: trypsin homology <TRY>
F;16-145,34-50,118-218,125-191,136-170,181-205/Disulfide bonds: #status predicted
F;49,93,185/Active site: His, Asp, Ser #status predicted
Query Match 83.2%; Score 79; DB 2; Length 231;
Best Local Similarity 75.0%; Pred. No. 1.2e-06;
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 IVGGYXCXHXSHQAHQVSLNS 20
| | | | | : | | | | |
Db 10 IVGGYECTKAYSQPHQVSLNS 29
| | | | | : | | | | |

RESULT 7
TRRT1
trypsin (EC 3.4.21.4) I precursor - rat
N;Alternate names: trypsinogen I
C;Species: Rattus norvegicus (Norway rat)
C;Date: 17-Dec-1982 #sequence_revision 17-Dec-1982 #text_change 24-Sep-1999
C;Accession: B22657; A00948
R;Crail, C.S.; Choo, Q.L.; Swift, G.H.; Quinto, C.; MacDonald, R.J.; Rutter, W.J.
J. Biol. Chem. 259, 14255-14264, 1984
A;Title: Structure of two related rat pancreatic trypsin genes.
A;Reference number: A22657; MUID:85054880; PMID:6094547
A;Accession: B22657
```



A;Cross-references: EMBL:X59012; NID:g57412; PIDN:CAA41751.1; PID:g57413  
A;Experimental source: pancreas  
C;Superfamily: trypsin; trypsin homology  
C;Keywords: hydrolase; protein digestion; serine proteinase; zymogen  
F;1-15/Domain: signal sequence #status predicted <SIG>  
F;16-24/Domain: activation peptide #status predicted <ACT>  
F;25-246/Product: trypsin V, a-form #status predicted <MAT>  
F;25-239/Domain: trypsin homology <TRY>  
F;31-160,49-65,133-233,140-206,171-185/Disulfide bonds: #status predicted  
F;64,108,200/Active site: His, Asp, Ser #status predicted  
Query Match 72.6%; Score 69; DB 2; Length 246;  
Best Local Similarity 65.0%; Pred. No. 8.2e-05;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
QY 1 IVGGYXCXHXSHQHQVSLNS 20  
Db 25 IVGGYTCQHSRNPVQVSLNA 44  
RESULT 12  
TRPTTR  
trypsin (EC 3.4.21.4) precursor - pig (tentative sequence)  
N;Contains: trypsinogen  
C;Species: Sus scrofa domestica (domestic pig)  
C;Date: 24-Apr-1984 #sequence revision 24-Apr-1984 #text\_change 31-Mar-2000  
C;Accession: A90641; A90368; A00947  
R;Charles M.; Rovey, M.; Guidoni, A.; Desnuelle, P.  
Biochim. Biophys. Acta 69, 115-129, 1963  
A;Title: Su le trypsinogene et la trypsine de porc.  
A;Reference number: A90641  
A;Accession: A90641  
A;Molecule type: protein  
A;Residues: 1-10 <CH>  
R;Hermanson, M.A.; Ericsson, L.H.; Neurath, H.; Walsh, K.A.  
Biochemistry 12, 3146-3153, 1973  
A;Title: Determination of the amino acid sequence of porcine trypsin by sequenator analysis  
A;Reference number: A90368; MUID:73258692; PMID:4738933  
A;Accession: A90368  
A;Molecule type: protein  
A;Residues: 9-231 <HR>  
A;Note: at position 20, Ile and Val occur alternatively  
C;Superfamily: trypsin; trypsin homology  
C;Keywords: hydrolase; pancreas; polymorphism; protein digestion; serine proteinase; zymogen  
F;1-231/Product: trypsinogen #status experimental <ZY>  
F;1-8/Domain: activation peptide #status experimental <APT>  
F;9-231/Product: trypsin #status experimental <MAT>  
F;9-224/Domain: trypsin homology <TRY>  
F;15-145,33-49,117-218,124-191,156-170,181-205/Disulfide bonds: #status predicted  
F;48,92,185/Active site: His, Asp, Ser #status predicted  
F;60,62,65,70/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted  
Query Match 69.5%; Score 66; DB 1; Length 231;  
Best Local Similarity 65.0%; Pred. No. 0.00027;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
QY 1 IVGGYXCXHXSHQHQVSLNS 20  
Db 9 IVGGYTCANSPVQVSLNS 28  
RESULT 13  
TRDGC  
trypsin (EC 3.4.21.4) precursor, cationic - dog  
N;Alternate names: cationic trypsinogen  
C;Species: Canis lupus familiaris (dog)  
C;Date: 30-Sep-1987 #sequence revision 30-Sep-1987 #text\_change 18-Jun-1999  
C;Accession: B26273  
R;Pinsky, S.D.; LaForge, K.S.; Scheele, G.  
Mol. Cell. Biol. 5, 2669-2676, 1985  
A;Title: Differential regulation of trypsinogen mRNA translation: full-length mRNA sequence  
A;Reference number: A26273; MUID:86284628; PMID:3841794  
A;Accession: B26273

A;Molecule type: mRNA  
A;Residues: 1-246 <PIN>  
A;Cross-references: GB:M11590; NID:g164096; PIDN:AAA30900.1; PID:g164097  
C;Superfamily: trypsin; trypsin homology  
C;Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen  
F;1-15/Domain: signal sequence #status predicted <SIG>  
F;16-23/Domain: activation peptide #status predicted <APT>  
F;24-246/Product: trypsin, cationic #status predicted <ENZ>  
F;24-239/Domain: trypsin homology <TRY>  
F;30-160,48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted  
F;63,107,200/Active site: His, Asp, Ser #status predicted  
F;75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted  
Query Match 69.5%; Score 66; DB 1; Length 246;  
Best Local Similarity 65.0%; Pred. No. 0.00028;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
QY 1 IVGGYXCXHXSHQHQVSLNS 20  
Db 24 IVGGYTCRNSRNPVQVSLNS 43  
RESULT 14  
TRRT2  
trypsin (EC 3.4.21.4) II precursor - rat  
N;Alternate names: trypsinogen II  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 05-Apr-1983 #sequence revision 30-Sep-1987 #text\_change 18-Jul-1997  
C;Accession: A22657; A00949  
R;Crak, C.S.; Choo, Q.L.; Swift, G.H.; Quinto, C.; MacDonald, R.J.; Rutter, W.J.  
J. Biol. Chem. 259, 14255-14264, 1984  
A;Title: Structure of two related rat pancreatic trypsin genes.  
A;Reference number: A22657; MUID:85054880; PMID:6094547  
A;Accession: A22657  
A;Molecule type: DNA  
A;Residues: 1-246 <CRA>  
R;MacDonald, R.J.; Stary, S.J.; Swift, G.H.  
J. Biol. Chem. 257, 9724-9732, 1982  
A;Title: Two similar but nonallelic rat pancreatic trypsinogens. Nucleotide sequences of  
A;Reference number: A00948; MUID:82365624; PMID:6896710  
A;Accession: A00949  
A;Molecule type: mRNA  
A;Residues: 9-246 <MAC>  
C;Comment: The trypsin II mRNA is present in much lower quantities than the trypsin I mRNA  
C;Genetics:  
A;Introns: 14/1; 67/2  
C;Superfamily: trypsin; trypsin homology  
C;Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen  
F;1-15/Domain: signal sequence #status predicted <SIG>  
F;16-23/Domain: activation peptide #status predicted <APT>  
F;24-246/Product: trypsin II #status predicted <ENZ>  
F;24-239/Domain: trypsin homology <TRY>  
F;30-160,48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted  
F;63,107,200/Active site: His, Asp, Ser #status predicted  
F;75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted  
Query Match 68.4%; Score 65; DB 1; Length 246;  
Best Local Similarity 65.0%; Pred. No. 0.00043;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
QY 1 IVGGYXCXHXSHQHQVSLNS 20  
Db 24 IVGGYTCQNSRNPVQVSLNS 43  
RESULT 15  
S12764  
trypsin (EC 3.4.21.4) III precursor - human  
N;Alternate names: Homo sapiens (man)  
C;Species: Homo sapiens (man)  
C;Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text\_change 22-Jun-1999  
C;Accession: S12764  
R;Tani, T.; Kawashima, I.; Mita, K.; Takiguchi, Y.  
Nucleic Acids Res. 18, 1631, 1990



A:Title: Nucleotide sequence of the human pancreatic trypsinogen III cDNA.  
A:Reference number: S12764; MUID:90221895; PMID:2326201  
A:Accession: S12764  
A:Molecule type: mRNA  
A:Residues: 1-247 <TAN>  
A:Cross-References: EMBL:X15505; NID:g37459; PIDN:CAA33527.1; PID:g37460  
C:Genetics:  
A:Gene: GDB:PRSS3; TRY3  
A:Cross-References: GDB:335297  
A:Map position: 7q35.7q35  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: calcium binding; hydrolase; pancreas; protein digestion; serine proteinase;  
F:1-15/Domain: signal sequence #status predicted <SIG>  
F:16-21/Domain: activation peptide #status predicted <APT>  
F:22-247/Product: trypsin III #status predicted <MAT>  
F:24-239/Domain: trypsin homology <TRY>  
F:30-160,48-64,139-206,171-185/Disulfide bonds: #status predicted  
F:63,107,200/Active site: His, Asp, Ser #status predicted  
F:75,77,80,85/Binding site: Calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 68.4%; Score 65; DB 2; Length 247;  
Best Local Similarity 65.0%; Pred. No. 0.00043;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 IVGGYXCXHSQAHOVSLS 20  
Db 24 IVGGYTCENSLPYQVLS 43

Search completed: February 12, 2003, 10:28:46  
Job time : 5.01493 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:17:50 ; Search time 5.01493 Seconds  
(without alignments)  
383.393 Million cell updates/sec

Title: US-10-036-371-2

Perfect score: 107

Sequence: 1 IVGGYTCGANTVPYQVSLNS 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 73.\*

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	229	1 TRBOTR	trypsin (EC 3.4.21
2	96	89.7	231	1 TRBOTR	trypsin (EC 3.4.21
3	92	86.0	246	1 TRPGC	trypsin (EC 3.4.21
4	90	84.1	245	1 TRRT2	trypsin (EC 3.4.21
5	89	83.2	247	2 S13813	trypsin (EC 3.4.21
6	87	81.3	247	1 TRDG	trypsin (EC 3.4.21
7	87	81.3	247	2 S12764	trypsin (EC 3.4.21
8	87	81.3	304	2 S33496	trypsin (EC 3.4.21
9	85	79.4	246	1 TRRT1	trypsin (EC 3.4.21
10	85	79.4	247	1 A25852	trypsin (EC 3.4.21
11	85	79.4	248	2 S55066	trypsin (EC 3.4.21
12	84	78.5	30	2 A61333	trypsin (EC 3.4.21
13	84	78.5	247	1 B25852	trypsin (EC 3.4.21
14	84	78.5	247	2 A27547	trypsin (EC 3.4.21
15	82	76.6	246	2 B25528	trypsin (EC 3.4.21
16	82	76.6	246	2 JQ1472	trypsin (EC 3.4.21
17	82	76.6	246	2 JQ1471	trypsin (EC 3.4.21
18	78.5	71.4	259	2 I38363	trypsin (EC 3.4.21
19	78	72.9	248	2 S55067	trypsin (EC 3.4.21
20	74	69.2	20	2 S50023	trypsin-like prote
21	74	69.2	40	2 S50021	trypsin-like prote
22	74	69.2	247	2 S05494	trypsin (EC 3.4.21
23	70	65.4	231	2 S31778	trypsin (EC 3.4.21
24	69	64.5	243	2 A35871	trypsin (EC 3.4.21
25	65	60.7	37	2 S03570	trypsin (EC 3.4.21
26	62	57.9	238	2 S31779	trypsin (EC 3.4.21
27	62	57.9	242	2 S31775	trypsin (EC 3.4.21
28	62	57.9	242	2 S31776	trypsin (EC 3.4.21
29	60	56.1	48	2 A61331	trypsin (EC 3.4.21

30	60	56.1	240	2 S39047	trypsin (EC 3.4.21
31	60	56.1	241	2 S39048	trypsin (EC 3.4.21
32	60	56.1	261	2 A31136	tissue kallikrein
33	59	55.1	229	1 TRDFS	trypsin (EC 3.4.21
34	58	54.2	25	2 A35545	T-kininogenase (EC
35	58	54.2	104	2 S15395	tissue kallikrein-
36	58	54.2	242	2 S15395	tissue kallikrein-
37	58	54.2	242	2 S49489	trypsin (EC 3.4.21
38	58	54.2	244	2 A44284	tissue kallikrein
39	57	53.3	37	2 D23863	tissue kallikrein
40	57	53.3	35	2 S23145	tissue kallikrein-
41	57	53.3	259	1 KQRTTN	tonin (EC 3.4.21-
42	57	53.3	265	1 KQRTT	tissue kallikrein
43	56	52.3	27	2 C61168	cocoonase (EC 3.4.
44	56	52.3	43	2 A61168	cocoonase (EC 3.4.
45	56	52.3	250	2 T01779	trypsin (EC 3.4.21
			250	2 S31384	trypsin (EC 3.4.21

#### ALIGNMENTS

##### RESULT 1

TRBOTR  
trypsin (EC 3.4.21.4) precursor - bovine  
N:Contains: trypsinogen  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 24-Apr-1984 #sequence\_revision 28-Feb-1986 #text\_change 18-Jul-1997  
C/Accession: A90164; A00946; S08774  
F/Mikes, O.; Holeysovsky, V.; Tomasek, V.; Sorm, F.  
Biochem. Biophys. Res. Commun. 24, 346-352, 1966  
A>Title: Covalent structure of bovine trypsinogen. The position of the remaining amides  
A/Reference number: A90164; MUID:67168848; PMID:5967094  
A/Accession: A90164  
A/Molecule type: protein  
A/Residues: 1-57, Q', 59-67, Q', 69-150, 'N', 152-176, 'N', 178-229 <MIK>  
R/Hartley, B.S.  
Philos. Trans. R. Soc. Lond. B257, 77-87, 1970  
A/Reference number: A93755  
R/Titani, K.; Ericsson, L.H.; Neurath, H.; Walsh, K.A.  
Biochemistry 14, 1358-1366, 1975  
A>Title: Amino acid sequence of dogfish trypsin.  
A/Reference number: A00950; MUID:75146445; PMID:1092332  
A/Contents: annotation; revisions  
A/Note: the sequence agrees with that shown  
R/Bode, W.; Schwager, P.  
J. Mol. Biol. 98, 693-717, 1975  
A/Title: The refined crystal structure of bovine beta-trypsin at 1.8 angstrom resolution  
A/Reference number: A92954; MUID:76072097; PMID:512  
A/Contents: annotation; X-ray crystallography; binding sites for calcium, substrate, and  
C/Comment: Trypsinogen is synthesized in the acinar cells of the pancreas.  
C/Comment: Autocatalytic cleavage after Lys-6 leads to beta-trypsin by releasing a termi  
S pseudotrypsin. A cleavage may also occur after Arg-105.  
C/Superfamily: trypsin; trypsin homology  
C/Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen  
F/1-229/Product: trypsinogen #status experimental <ZYM>  
F/1-6/Domain: activation peptide #status experimental <APT>  
F/7-222/Domain: trypsin homology <TRY>  
F/7-131,132-229/Product: alpha-trypsin #status experimental <MPT>  
F/6-7/Cleavage site: Lys-116 (enteropeptidase) #status experimental  
F/13-143,31-47,115-216,122-189,154-168,179-203/Disulfide bonds: #status experimental  
F/46,90,183/Active site: His, Asp, Ser #status experimental  
F/58,60,63,68/Binding site: calcium (Glu, Asn, Val, Glu) #status experimental  
F/131-132/Cleavage site: Lys-Ser (autolytic) #status experimental

Query Match 100.0%; Score 107; DB 1; Length 229;  
Best Local Similarity 100.0%; Pred. No. 6.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 3; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20

|||||

Db 7 IVGGYTCGANTVPYQVSLNS 26

```
Db 24 IVGGYTCGRNSVPYQVSLNS 43

RESULT 4
TRFCTR
trypsin (EC 3.4.21.4) precursor - pig (tentative sequence)
N/Contains: trypsinogen
C/Species: Sus scrofa domestica (domestic pig)
C/Date: 24-Apr-1984 #sequence_revision 24-Apr-1984 #text_change 31-Mar-2000
C/Accession: A90641; A90368; A00947
R/Charles, M.; Rovey, M.; Guidoni, A.; Desnuelle, P.
Biochim. Biophys. Acta 69, 115-129, 1963
A/Title: Su le trypsinogene et la trypsine de porc.
A/Reference number: A90641
A/Accession: A90641
A/Molecule type: protein
A/Residues: 1-10 <CRA>
R/Hermanson, M.A.; Ericsson, L.H.; Neurath, H.; Walsh, K.A.
Biochemistry 12, 3146-3153, 1973
A/Title: Determination of the amino acid sequence of porcine trypsin by sequenator analysis
A/Reference number: A90368; MUID:73258692; PMID:4738933
A/Accession: A90368
A/Molecule type: protein
A/Residues: 9-231 <HER>
A/Note: at position 20, Ile and Val occur alternatively
C/Superfamily: trypsin; trypsin homology
C/Keywords: hydrolase; pancreas; polymorphism; protein digestion; serine proteinase; zymogen
F/1-231/Product: trypsinogen #status experimental <ZYM>
F/1-8/Domain: activation peptide #status experimental <WAT>
F/9-231/Product: trypsin #status experimental <WAT>
F/9-224/Domain: trypsin homology <TRY>
F/15-145,33-49,117-218,124-191,156-170,181-205/Disulfide bonds: #status predicted
F/15-145,33-49,117-218,124-191,156-170,181-205/Disulfide bonds: #status predicted
F/48,92,185/Active site: His, Asp, Ser #status predicted
F/60,62,65,70/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 89.7%; Score 96; DB 1; Length 231;
Best Local Similarity 85.0%; Pred. No. 3.1e-07;
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IVGGYTCGANTVPYQVSLNS 20
Db 9 IVGGYTCANSPYQVSLNS 28

RESULT 3
TRDGC
trypsin (EC 3.4.21.4) precursor, cationic - dog
N/Alternate names: cationic trypsinogen
C/Species: Canis lupus familiaris (dog)
C/Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 18-Jun-1999
C/Accession: B26273
R/Pinsky, S.D.; LaForge, K.S.; Scheele, G.
Mol. Cell. Biol. 5, 2669-2676, 1985
A/Title: Differential regulation of trypsinogen mRNA translation: full-length mRNA sequence
A/Reference number: A26273; MUID:86284628; PMID:3841794
A/Accession: B26273
A/Molecule type: mRNA
A/Residues: 1-246 <PIN>
A/Cross-references: GB:M1590; NID:G164096; PIDN:AAA30900.1; PID:G164097
C/Superfamily: trypsin; trypsin homology
C/Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen
F/1-15/Domain: signal sequence #status predicted <SIG>
F/16-23/Domain: activation peptide #status predicted <APT>
F/24-246/Product: trypsin, cationic #status predicted <ENZ>
F/24-239/Domain: trypsin homology <TRY>
F/30-160,48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted
F/63,107,200/Active site: His, Asp, Ser #status predicted
F/75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 86.0%; Score 92; DB 1; Length 246;
Best Local Similarity 85.0%; Pred. No. 1.4e-06;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IVGGYTCGANTVPYQVSLNS 20
Db 1 IVGGYTCANSPYQVSLNS 20

RESULT 5
S13813
trypsin (EC 3.4.21.4) - bovine
C/Species: Bos primigenius taurus (cattle)
C/Date: 02-Dec-1993 #sequence_revision 03-Aug-1995 #text_change 22-Jun-1999
C/Accession: S13813
R/le Huerou, I.; Wicker, C.; Guilloteau, P.; Toullec, R.; Puigserver, A.
Eur. J. Biochem. 193, 767-773, 1990
A/Title: Isolation and nucleotide sequence of cDNA clone for bovine pancreatic anionic trypsin
A/Reference number: S13813; MUID:91065383; PMID:1701147
A/Accession: S13813
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-247 <HUE>
A/Cross-references: EMBL:X54703; NID:G829; PIDN:CAA38513.1; PID:G830
C/Superfamily: trypsin; trypsin homology
C/Keywords: hydrolase; protein digestion; serine proteinase
F/24-239/Domain: trypsin homology <TRY>
F/63,107,200/Active site: His, Asp, Ser #status predicted

Query Match 83.2%; Score 89; DB 2; Length 247;
Best Local Similarity 80.0%; Pred. No. 3.9e-06;
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IVGGYTCGANTVPYQVSLNS 20
Db 24 IVGGYTCANSPYQVSLNS 43
```

```
Db 24 IVGGYTCGRNSVPYQVSLNS 43

RESULT 4
TRRT2
trypsin (EC 3.4.21.4) II precursor - rat
N/Alternate names: trypsinogen II
C/Species: Rattus norvegicus (Norway rat)
C/Date: 05-Apr-1983 #sequence_revision 30-Sep-1987 #text_change 18-Jul-1997
C/Accession: A22657; A00949
R/Craik, C.S.; Choo, Q.L.; Swift, G.H.; Quinto, C.; MacDonald, R.J.; Rutter, W.J.
J. Biol. Chem. 259, 14255-14264, 1984
A/Title: Structure of two related rat pancreatic trypsin genes.
A/Reference number: A22657; MUID:85054880; PMID:6094547
A/Accession: A22657
A/Molecule type: DNA
A/Residues: 1-246 <CRA>
R/MacDonald, R.J.; Stary, S.J.; Swift, G.H.
J. Biol. Chem. 257, 9724-9732, 1982
A/Title: Two similar but nonallelic rat pancreatic trypsinogens. Nucleotide sequences of
A/Reference number: A00948; MUID:82265624; PMID:6896710
A/Accession: A00949
A/Molecule type: mRNA
A/Residues: 9-246 <MAC>
C/Comment: The trypsin II mRNA is present in much lower quantities than the trypsin I mRNA
C/Genetics:
A/Introns: 14/1; 67/2
C/Superfamily: trypsin; trypsin homology
C/Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen
F/1-15/Domain: signal sequence #status predicted <SIG>
F/16-23/Domain: activation peptide #status predicted <APT>
F/24-246/Product: trypsin II #status predicted <ENZ>
F/24-239/Domain: trypsin homology <TRY>
F/30-160,48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted
F/63,107,200/Active site: His, Asp, Ser #status predicted
F/75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 84.1%; Score 90; DB 1; Length 246;
Best Local Similarity 85.0%; Pred. No. 2.8e-06;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IVGGYTCGANTVPYQVSLNS 20
Db 24 IVGGYTCGRNSVPYQVSLNS 43

RESULT 5
S13813
trypsin (EC 3.4.21.4) - bovine
C/Species: Bos primigenius taurus (cattle)
C/Date: 02-Dec-1993 #sequence_revision 03-Aug-1995 #text_change 22-Jun-1999
C/Accession: S13813
R/le Huerou, I.; Wicker, C.; Guilloteau, P.; Toullec, R.; Puigserver, A.
Eur. J. Biochem. 193, 767-773, 1990
A/Title: Isolation and nucleotide sequence of cDNA clone for bovine pancreatic anionic trypsin
A/Reference number: S13813; MUID:91065383; PMID:1701147
A/Accession: S13813
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-247 <HUE>
A/Cross-references: EMBL:X54703; NID:G829; PIDN:CAA38513.1; PID:G830
C/Superfamily: trypsin; trypsin homology
C/Keywords: hydrolase; protein digestion; serine proteinase
F/24-239/Domain: trypsin homology <TRY>
F/63,107,200/Active site: His, Asp, Ser #status predicted

Query Match 83.2%; Score 89; DB 2; Length 247;
Best Local Similarity 80.0%; Pred. No. 3.9e-06;
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IVGGYTCGANTVPYQVSLNS 20
Db 24 IVGGYTCGRNSVPYQVSLNS 43
```

## RESULT 6

TRQD  
 trypsin (EC 3.4.21.4) precursor, anionic - dog  
 N/Alternate names: cationic trypsinogen  
 C/Species: Canis lupus familiaris (dog)  
 C/Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 18-Jun-1999  
 C/Accession: A26273  
 R/Pinsky, S.D.; LaForge, K.S.; Scheele, G.  
 Mol. Cell. Biol. 5, 2669-2676, 1985  
 A/Title: Differential regulation of trypsinogen mRNA translation: full-length mRNA sequence  
 A/Reference number: A26273; MUID:86284628; PMID:3841794  
 A/Accession: A26273  
 A/Molecule type: mRNA  
 A/Residues: 1-247 <PIN>  
 A/Cross-references: GB:M11589; NID:g164094; PIDN:AAA30899.1; PID:g164095  
 C/Superfamily: trypsin; trypsin homology  
 C/Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen  
 F:1-15/Domain: signal sequence #status predicted <SIG>  
 F:16-23/Domain: activation peptide #status predicted <APT>  
 F:24-247/Product: trypsin, anionic #status predicted <ENZ>  
 F:24-239/Domain: trypsin homology <TRY>  
 F:30-160,48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted  
 F:63,107,200/Active site: His, Asp, Ser #status predicted  
 F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 81.3%; Score 87; DB 1; Length 247;  
 Best Local Similarity 80.0%; Pred. No. 7.9e-06;  
 Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20

||||| :|||||

Db 24 IVGGYTCENSLPYQVSLNA 43

## RESULT 7

S12764  
 trypsin (EC 3.4.21.4) III precursor - human  
 C/Species: Homo sapiens (man)  
 C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 22-Jun-1999  
 C/Accession: S12764  
 R/Tani, T.; Kawashima, I.; Mita, K.; Takiguchi, Y.  
 Nucleic Acids Res. 18, 1631, 1990  
 A/Title: Nucleotide sequence of the human pancreatic trypsinogen III cDNA.  
 A/Reference number: S12764; MUID:90221895; PMID:2326201  
 A/Accession: S12764  
 A/Molecule type: mRNA  
 A/Residues: 1-247 <TAN>  
 A/Cross-references: EMBL:X15505; NID:g37459; PIDN:CAA33527.1; PID:g37460  
 C/Genetics:  
 A/Gene: GDB:PRSS3; TRY3  
 A/Cross-references: GDB:335297  
 A/Map position: 7q35-7q35  
 C/Superfamily: trypsin; trypsin homology  
 C/Keywords: calcium binding; hydrolase; pancreas; protein digestion; serine proteinase;  
 F:1-15/Domain: signal sequence #status predicted <SIG>  
 F:16-21/Domain: activation peptide #status predicted <APT>  
 F:22-247/Product: trypsin III #status predicted <MAT>  
 F:24-239/Domain: trypsin homology <TRY>  
 F:30-160,48-64,139-206,171-185/Disulfide bonds: #status predicted  
 F:63,107,200/Active site: His, Asp, Ser #status predicted  
 F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 81.3%; Score 87; DB 2; Length 247;  
 Best Local Similarity 80.0%; Pred. No. 7.9e-06;  
 Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20

||||| :|||||

Db 24 IVGGYTCENSLPYQVSLNS 43

## RESULT 8

S33496  
 trypsin (EC 3.4.21.4) IV form a - human  
 C/Species: Homo sapiens (man)  
 C/Date: 03-Mar-1994 #sequence\_revision 03-Aug-1995 #text\_change 15-Aug-1997  
 C/Accession: S33496  
 R/Wiegand, U.; Corbach, S.; Minn, A.; Kang, J.; Mueller-Hill, B.  
 submitted to the EMBL Data Library, March 1993  
 A/Description: Identification, cloning and characterization of a cDNA encoding a human b  
 A/Reference number: S33496  
 A/Accession: S33496  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-304 <WIE>  
 A/Cross-references: EMBL:X72781  
 C/Genetics:  
 A/Gene: GDB:PRSS4; TRY4  
 A/Cross-references: GDB:335300  
 A/Map position: 7q35-7q35  
 C/Superfamily: trypsin; trypsin homology  
 C/Keywords: hydrolase; serine proteinase  
 F:81-236/Domain: trypsin homology <TRY>  
 F:120,164,257/Active site: His, Asp, Ser #status predicted

Query Match 81.3%; Score 87; DB 2; Length 304;  
 Best Local Similarity 80.0%; Pred. No. 9.7e-06;  
 Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20

||||| :|||||

Db 81 IVGGYTCENSLPYQVSLNS 100

## RESULT 9

TRKII  
 trypsin (EC 3.4.21.4) I precursor - rat  
 N/Alternate names: trypsinogen I  
 C/Species: Rattus norvegicus (Norway rat)  
 C/Date: 17-Dec-1982 #sequence\_revision 17-Dec-1982 #text\_change 24-Sep-1999  
 C/Accession: B22657; A00948  
 R/Craik, C.S.; Choo, Q.L.; Swift, G.H.; Quinto, C.; MacDonald, R.J.; Rutter, W.J.  
 J. Biol. Chem. 259, 14255-14264, 1984  
 A/Title: Structure of two related rat pancreatic trypsin genes.  
 A/Reference number: A22657; MUID:85054880; PMID:6094547  
 A/Accession: B22657  
 A/Molecule type: DNA  
 A/Residues: 1-246 <CRA>  
 A/Cross-references: GB:J00778; NID:g206507; PIDN:AAA98518.1; PID:g206508  
 A/Note: the authors translated the codon ATC for residue 6 as Leu and GAC for residue 17  
 R/MacDonald, R.J.; Stary, S.J.; Swift, G.H.  
 J. Biol. Chem. 257, 9724-9732, 1982  
 A/Title: Two similar but nonallelic rat pancreatic trypsinogens. Nucleotide sequences of  
 A/Reference number: A00948; MUID:82265624; PMID:6896710

A/Accession: A00948  
 A/Molecule type: mRNA  
 A/Residues: 1-246 <MAC>  
 A/Cross-references: GB:J00778; NID:g206507; PIDN:AAA98518.1; PID:g206508  
 C/Genetics:

A/Introns: 14/1; 67/2; 152/1; 197/3  
 C/Superfamily: trypsin; trypsin homology  
 C/Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen  
 F:1-15/Domain: signal sequence #status predicted <SIG>  
 F:16-23/Domain: activation peptide #status predicted <APT>  
 F:24-246/Product: trypsin I #status predicted <ENZ>  
 F:24-239/Domain: trypsin homology <TRY>  
 F:30-160,48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted  
 F:63,107,200/Active site: His, Asp, Ser #status predicted  
 F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 79.4%; Score 85; DB 1; Length 246;  
 Best Local Similarity 80.0%; Pred. No. 1.6e-05;  
 Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20  
 ||||| :|||||  
 Db 24 IVGGYTCPEHSVPYQVSLNS 43

## RESULT 10

A25852  
 trypsin (EC 3.4.21.4) I precursor [validated] - human  
 N/Alternate names: trypsin, cationic; trypsinogen I  
 C/Species: Homo sapiens (man)  
 C/Date: 03-Mar-1994 #sequence\_revision 03-Mar-1994 #text\_change 08-Dec-2000  
 C/Accession: A25852; B61066; A43988  
 R/Emi, M.; Nakamura, Y.; Ogawa, M.; Yamamoto, T.; Nishide, T.; Mori, T.; Matsubara, K.  
 Gene 41, 305-310, 1986  
 A/Title: Cloning, characterization and nucleotide sequences of two cDNAs encoding human  
 A/Reference number: A91544; MUID:86221712; PMID:3011602  
 A/Accession: A25852  
 A/Molecule type: mRNA  
 A/Residues: 1-247 <EMI>  
 A/Cross-references: GB:M22612; NID:G521215; PIDN:AAA61231.1; PID:G521216  
 R/Kimland, M.; Russick, C.; Marks, W.H.; Borgetroem, A.  
 Clin. Chim. Acta 184, 31-46, 1989  
 A/Title: Immunoreactive anionic and cationic trypsin in human serum.  
 A/Reference number: A61066; MUID:90091010; PMID:2598466  
 A/Accession: B61066  
 A/Molecule type: protein  
 A/Residues: 16-43 <KIM>  
 R/Koivunen, E.; Huhtala, M.L.; Stenman, U.H.  
 J. Biol. Chem. 264, 14095-14099, 1989  
 A/Title: Human ovarian tumor-associated trypsin. Its purification and characterization  
 A/Reference number: A43988; MUID:89340515; PMID:2503510  
 A/Accession: A43988  
 A/Molecule type: protein  
 A/Residues: 16-54 <KOI>  
 A/Experimental source: mucinous ovarian tumor cyst fluid  
 C/Genetic8:  
 A/Gene: GDB:PRSS1; TRY1  
 A/Cross-references: GDB:I19620; OMIM:276000  
 A/Map position: 7q35-7q35  
 A/Note: The human genome contains at least ten trypsin genes or pseudogenes, at least two  
 C/Superfamily: trypsin; trypsin homology  
 C/Keywords: hydrolase; pancreas; phosphoprotein; protein digestion; serine proteinase; z  
 F/1-15/Domain: signal sequence #status predicted <SIG>  
 F/16-246/Product: trypsinogen I #status experimental <ZYM>  
 F/16-23/Domain: activation peptide #status experimental <APT>  
 F/24-246/Product: trypsin I #status predicted <ENZ>  
 F/24-239/Domain: trypsin homology <TRY>  
 F/30-160,48-64,139-206,171-185,196-220/Disulfide bonds: #status predicted  
 F/63,107,200/Active site: His, Asp, Ser #status predicted  
 F/75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 79.4%; Score 85; DB 1; Length 247;  
 Best Local Similarity 80.0%; Pred. No. 1.6e-05;  
 Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20  
 ||||| :|||||  
 Db 24 IVGGYCNCEHSVPYQVSLNS 43

## RESULT 11

S55066  
 trypsin (EC 3.4.21.4) II precursor, pancreatic (clone 2-P29) - chicken  
 N/Alternate names: trypsinogen II  
 C/Species: Gallus gallus (chicken)  
 C/Date: 23-Aug-1995 #sequence\_revision 19-Oct-1995 #text\_change 22-Jun-1999  
 C/Accession: S55066; S72347  
 R/Wang, K.; Gan, L.; Lee, I.; Hood, L.  
 Biochem. J. 307, 471-479, 1995  
 A/Title: Isolation and characterization of the chicken trypsinogen gene family.  
 A/Reference number: S55065; MUID:95251611; PMID:7733885  
 A/Accession: S55066  
 A/Molecule type: mRNA

A/Residues: 1-248 <WAN1>  
 A/Cross-references: EMBL:U15157; NID:G603906; PIDN:AAA79914.1; PID:G603907  
 A/Experimental source: clone 2-P29  
 A/Accession: S72347  
 A/Molecule type: DNA  
 A/Residues: 1-248 <WAN2>  
 A/Cross-references: EMBL:U15157; NID:G603906; PIDN:AAA79914.1; PID:G603907  
 A/Experimental source: clone 2-P29  
 C/Superfamily: trypsin; trypsin homology  
 C/Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen  
 F/1-16/Domain: signal sequence #status predicted <SIG>  
 F/17-25/Domain: activation peptide #status predicted <APT>  
 F/26-248/Product: trypsin II #status predicted <MAT>  
 F/26-241/Domain: trypsin homology <TRY>  
 F/65,109,202/Active site: His, Asp, Ser #status predicted

Query Match 79.4%; Score 85; DB 2; Length 248;  
 Best Local Similarity 80.0%; Pred. No. 1.6e-05;  
 Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20  
 ||||| :|||||  
 Db 26 IVGGYTCPEHSVPYQVSLNS 45

## RESULT 12

A61333  
 trypsin (EC 3.4.21.4) - edible frog (fragment)  
 C/Species: Rana esculenta (edible frog)  
 C/Date: 17-Jul-1994 #sequence\_revision 17-Jul-1994 #text\_change 07-May-1999  
 C/Accession: A61333  
 R/Pies, W.; Zwilling, R.; Woodbury, R.G.; Neurath, H.  
 FEBS Lett. 109, 45-49, 1980  
 A/Title: Amino-terminal amino acid sequences and the evolution of frog (Rana esculenta)  
 A/Reference number: A61333; MUID:80113255; PMID:6965480  
 A/Accession: A61333  
 A/Molecule type: protein  
 A/Residues: 1-30 <PIE>  
 C/Superfamily: trypsin; trypsin homology  
 C/Keywords: hydrolase; protein digestion; serine proteinase

Query Match 78.5%; Score 84; DB 2; Length 30;  
 Best Local Similarity 75.0%; Pred. No. 3e-06;  
 Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20  
 ||||| :|||||  
 Db 1 IVGGFTCSKNSVPYQASLNS 20

## RESULT 13

B25852  
 trypsin (EC 3.4.21.4) II precursor [validated] - human  
 N/Alternate names: trypsin 2; trypsin, anionic; trypsinogen II  
 C/Species: Homo sapiens (man)  
 C/Date: 03-Mar-1994 #sequence\_revision 03-Mar-1994 #text\_change 08-Dec-2000  
 C/Accession: B25852; A61066; B43988  
 R/Emi, M.; Nakamura, Y.; Ogawa, M.; Yamamoto, T.; Nishide, T.; Mori, T.; Matsubara, K.  
 Gene 41, 305-310, 1986  
 A/Title: Cloning, characterization and nucleotide sequences of two cDNAs encoding human I  
 A/Reference number: A91544; MUID:86221712; PMID:3011602  
 A/Accession: B25852  
 A/Molecule type: mRNA  
 A/Residues: 1-247 <EMI>  
 A/Cross-references: GB:M27602; NID:G521217; PIDN:AAA61232.1; PID:G521218  
 R/Kimland, M.; Russick, C.; Marks, W.H.; Borgetroem, A.  
 Clin. Chim. Acta 184, 31-46, 1989  
 A/Title: Immunoreactive anionic and cationic trypsin in human serum.  
 A/Reference number: A61066; MUID:90091010; PMID:2598466  
 A/Accession: A61066  
 A/Molecule type: protein  
 A/Residues: 16-39, X', 41-42, 'XXXX', 47-49 <KIM>  
 R/Koivunen, E.; Huhtala, M.L.; Stenman, U.H.

J. Biol. Chem. 264, 14095-14099, 1989  
A:Title: Human ovarian tumor-associated trypsin. Its purification and characterization  
A:Reference number: A43988; MUID:89340515; PMID:2503510  
A:Accession: B43988  
A:Molecule type: protein  
A:Residues: 16-49 <KOI>  
A:Experimental source: mucinous ovarian tumor cyst fluid  
C:Genetics:  
A:Gene: GDB:PRSS2; TRY2  
A:Cross-references: GDB:335289; OMIM:601564  
A:Map position: 7q35-7q35  
A:Note: the human genome contains at least ten trypsin genes or pseudogenes, at least two of which are expressed in the ovary.  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen  
F:1-15/Domain: signal sequence #status predicted <SIG>  
F:16-246/Product: trypsinogen II #status experimental <APT>  
F:16-23/Domain: activation peptide #status predicted <ENZ>  
F:24-246/Product: trypsin II #status predicted <TRY>  
F:24-239/Domain: trypsin homology <TRY>  
F:30-160,48-64,171-185,196-220/Disulfide bonds: #status predicted  
F:63,107,200/Active site: His, Asp, Ser #status predicted  
F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted  
Query Match 78.5%; Score 84; DB 1; Length 247;  
Best Local Similarity 80.0%; Pred. No. 2.3e-05;  
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 IVGGYTCGANTVPYQVSLNS 20  
Db 24 IVGGYICEENSVPYQVSLNS 43  
RESULT 14  
A27547  
trypsin (EC 3.4.21.4) precursor, cationic - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 22-Jun-1999  
C:Accession: A27547  
R:Fletcher, T.S.; Alhadeff, M.; Craik, C.S.; Largman, C.  
Biochemistry 26, 3081-3086, 1987  
A:Title: Isolation and characterization of a cDNA encoding rat cationic trypsinogen.  
A:Reference number: A27547; MUID:87271609; PMID:3607011  
A:Accession: A27547  
A:Molecule type: mRNA  
A:Residues: 1-247 <ELE>  
A:Cross-references: GB:M16624; NID:g206498; PIDN:AAA41985.1; PID:g206499  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: calcium binding; hydrolase; protein digestion; serine proteinase  
F:25-240/Domain: trypsin homology <TRY>  
F:31-161,49-65,133-234,140-207,172-186/Disulfide bonds: #status predicted  
F:64,108,201/Active site: His, Asp, Ser #status predicted  
F:76,78,81,86/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted  
Query Match 78.5%; Score 84; DB 2; Length 247;  
Best Local Similarity 75.0%; Pred. No. 2.3e-05;  
Matches 15; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
QY 1 IVGGYTCGANTVPYQVSLNS 20  
Db 25 IVGGYTCQNSLPYQVSLNA 44  
RESULT 15  
B25528  
trypsin (EC 3.4.21.4) precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 22-Jun-1999  
C:Accession: B25528  
R:Stevenson, B.J.; Hagenbuechle, O.; Wellauer, P.K.  
Nucleic Acids Res. 14, 8307-8330, 1986  
A:Title: Sequence organisation and transcriptional regulation of the mouse elastase II gene.  
A:Reference number: A93646; MUID:87066713; PMID:3641189  
A:Accession: B25528

A:Molecule type: mRNA  
A:Residues: 1-246 <STE>  
A:Cross-references: GB:X04574; NID:g54918; PIDN:CAA28243.1; PID:g54919  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: calcium binding; hydrolase; protein digestion; serine proteinase  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-246/Product: trypsin #status predicted <MAT>  
F:24-239/Domain: trypsin homology <TRY>  
F:30-160,48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted  
F:63,107,200/Active site: His, Asp, Ser #status predicted  
F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted  
Query Match 76.6%; Score 82; DB 2; Length 246;  
Best Local Similarity 75.0%; Pred. No. 4.6e-05;  
Matches 15; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
QY 1 IVGGYTCGANTVPYQVSLNS 20  
Db 24 IVGGYTCRESSVPYQVSLNA 43  
Search completed: February 12, 2003, 10:28:47  
Job time : 6.01493 secs

THIS PAGE BLANK (USPTO)



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:04:45 ; Search time 2.68657 Seconds  
(without alignments)  
308.768 Million cell updates/sec

Title: US-10-036-371-2

Perfect score: 107

Sequence: 1 IVGGYTCGANTVPYQVSLNS 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt 40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	100.0	243	1	TRY1_BOVIN
2	96	89.7	231	1	TRY1_PIG
3	92	86.0	246	1	TRY1_CANFA
4	90	84.1	246	1	TRY2_RAT
5	89	83.2	247	1	TRY2_BOVIN
6	87	81.3	247	1	TRY2_CANFA
7	87	81.3	247	1	TRY3_HUMAN
8	87	81.3	304	1	TRY1_RAT
9	85	79.4	247	1	TRY4_HUMAN
10	85	79.4	247	1	TRY1_HUMAN
11	85	79.4	248	1	TRY3_CHICK
12	84	78.5	244	1	TRY2_XENLA
13	84	78.5	247	1	TRY2_HUMAN
14	84	78.5	247	1	TRY3_RAT
15	82	76.6	246	1	TRY2_MOUSE
16	82	76.6	246	1	TRY4_RAT
17	82	76.6	246	1	TRYB_RAT
18	78	72.9	248	1	TRY1_CHICK
19	78	72.9	248	1	TRY2_CHICK
20	74	69.2	247	1	TRY4_RAT
21	70	65.4	231	1	TRY2_SALSA
22	69	64.5	243	1	TRY1_XENLA
23	62	57.9	238	1	TRY3_SALSA
24	62	57.9	242	1	TRY1_SALSA
25	60	56.1	241	1	TRY1_GADMO
26	60	56.1	241	1	TRY3_GADMO
27	60	56.1	261	1	KLK7_RAT
28	59	55.1	229	1	TRY7_SQAC
29	58	54.2	244	1	KLKA_RAT
30	58	54.2	259	1	KLK9_RAT
31	57	53.3	259	1	KLK2_RAT
32	57	53.3	261	1	KLK1_RAT
33	56	52.3	14	1	COCO_LIMPO

## RESULT 1

ID	TRY1_BOVIN	STANDARD	PRT	243 AA
AC	P00760;			
DT	21-JUL-1986	(Rel. 01, Created)		
DT	01-NOV-1997	(Rel. 35, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	Trypsinogen, cationic precursor (EC 3.4.21.4)	(Beta-trypsin)		
DE	(Fragment)			
OS	Bos taurus (Bovine)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Pancreas;			
RA	Okajima T., Maniwa M., Nagao S., Fujikawa H., Goto S.;			
RL	Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE OF 15-243, AND DISULFIDE BONDS.			
RX	MEDLINE=67168848; PubMed=5967094;			
RA	Mikes O., Holesovsky V., Tomasek V., Sorm F.;			
RT	"Covalent structure of bovine trypsinogen. The position of the remaining amides.";			
RL	Biochem. Biophys. Res. Commun. 24:346-352(1966).			
RN	[3]			
RP	REVISIONS.			
RX	MEDLINE=72035053; PubMed=4399051;			
RA	Hartley B.S.;			
RT	"Homologies in serine proteinases.";			
RL	Philos. Trans. R. Soc. Lond., B, Biol. Sci. 257:77-87(1970).			
RN	[4]			
RP	REVISIONS.			
RX	MEDLINE=75146445; PubMed=1092332;			
RA	Titani K., Ericsson L.H., Neurath H., Walsh K.A.;			
RT	"Amino acid sequence of dogfish trypsin.";			
RL	Biochemistry 14:1358-1366(1975).			
RN	[5]			
RP	X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF CALCIUM-BINDING SITE.			
RX	MEDLINE=76072097; PubMed=512;			
RA	Bode W., Schwager P.;			
RT	"The refined crystal structure of bovine beta-trypsin at 1.8-A resolution. II. Crystallographic refinement, calcium binding site, benzamide binding site and active site at pH 7.0.";			
RL	J. Mol. Biol. 98:693-717(1975).			
RN	[6]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=77112431; PubMed=556951;			
RA	Kossiakoff A.A., Chambers J.L., Kay L.M., Stroud R.M.;			
RT	"Structure of bovine trypsinogen at 1.9-A resolution.";			
RL	Biochemistry 16:654-664(1977).			
RN	[7]			
RP	DISULFIDE BONDS.			
RX	MEDLINE=66079271; PubMed=5892911;			
RA	Kauffman D.L.;			

34	56	52.3	250	1	TRYP_PLEPL	P35034 pleuronecte
35	56	52.3	259	1	KLKC_RAT	P36376 rattus norv
36	55	51.4	280	1	TRY2_DROME	P42280 drosophila
37	54	50.5	261	1	KLK1_MOUSE	P15947 mus musculu
38	53	49.5	256	1	TRYE_DROER	P54627 drosophila
39	53	49.5	256	1	TRYE_DROME	P35005 drosophila
40	53	49.5	261	1	KLK8_MOUSE	P07628 mus musculu
41	53	49.5	261	1	KLK8_RAT	P36374 rattus norv
42	53	49.5	261	1	KLKD_MOUSE	P36368 mus musculu
43	53	49.5	261	1	KLKG_MOUSE	P04071 mus musculu
44	53	49.5	261	1	KLKZ_MOUSE	P36369 mus musculu
45	52	48.6	259	1	DEF3_DERFA	P49275 dermatophag

## ALIGNMENTS

RT "The disulphide bridges of trypsin.";  
RL J. Mol. Biol. 12:929-932(1965).  
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
CC -1- SUBCELLULAR LOCATION: Extracellular.  
CC -1- TISSUE SPECIFICITY: SYNTHESIZED IN THE ACINAR CELLS OF THE  
CC PANCREAS.  
CC -1- PTM: AUTOCATALYTIC CLEAVAGE AFTER LYS-20 LEADS TO BETA-TRYPSIN BY  
CC RELEASING A TERMINAL HEXAPEPTIDE. SUBSEQUENT CLEAVAGE AFTER  
CC LYS-145 LEADS TO ALPHA-TRYPSIN. FURTHER CLEAVAGE AFTER LYS-190  
CC YIELDS PSEUDOTRYPSIN. A CLEAVAGE MAY ALSO OCCUR AFTER ARG-119.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
CC -1- DATABASE: NAME=worthington enzyme manual;  
CC WWW="http://www.worthington-biochem.com/manual/t/TRY.html".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR ENBL; D38507; BAA07516.1; -;  
DR PIR; A00946; TRBOTR.  
DR PDB; 2TGA; 09-APR-85.  
DR PDB; 1TGB; 14-MAR-85.  
DR PDB; 1TGC; 09-APR-85.  
DR PDB; 2TGD; 07-MAY-86.  
DR PDB; 1TGN; 22-OCT-84.  
DR PDB; 2TGP; 14-MAR-85.  
DR PDB; 1TGS; 14-MAR-85.  
DR PDB; 1TGT; 15-OCT-90.  
DR PDB; 2TGT; 09-APR-85.  
DR PDB; 1TLD; 15-OCT-92.  
DR PDB; 2TLD; 15-JUL-92.  
DR PDB; 1TPA; 14-MAR-85.  
DR PDB; 2TPI; 15-JUL-91.  
DR PDB; 3TPI; 14-MAR-85.  
DR PDB; 4TPI; 08-NOV-85.  
DR PDB; 1TPO; 14-MAR-85.  
DR PDB; 1TPR; 16-APR-87.  
DR PDB; 1NTP; 15-APR-91.  
DR PDB; 3PTB; 14-MAR-85.  
DR PDB; 2PTC; 14-MAR-85.  
DR PDB; 2PTN; 15-APR-91.  
DR PDB; 3PTN; 09-APR-85.  
DR PDB; 5PTP; 07-JUL-97.  
DR PDB; 1TAB; 15-JUL-92.  
DR PDB; 1GBT; 31-JAN-94.  
DR PDB; 1PPC; 31-JAN-94.  
DR PDB; 1PPE; 31-JAN-94.  
DR PDB; 1PPH; 31-JAN-94.  
DR PDB; 1SME; 31-JUL-94.  
DR PDB; 1TNG; 30-NOV-94.  
DR PDB; 1TNH; 30-NOV-94.  
DR PDB; 1TNI; 30-NOV-94.  
DR PDB; 1TNJ; 30-NOV-94.  
DR PDB; 1TNK; 30-NOV-94.  
DR PDB; 1TNL; 30-NOV-94.  
DR PDB; 1BTP; 29-JAN-96.  
DR PDB; 1BTW; 15-OCT-95.  
DR PDB; 1BTX; 15-OCT-95.  
DR PDB; 1BTY; 15-OCT-95.  
DR PDB; 1BTZ; 15-OCT-95.  
DR PDB; 1JRS; 14-OCT-96.  
DR PDB; 1JRT; 14-OCT-96.  
DR PDB; 1MAX; 14-OCT-96.  
DR PDB; 1MAY; 14-OCT-96.  
DR PDB; 1MTS; 20-AUG-97.  
DR PDB; 1MTU; 12-NOV-97.  
DR PDB; 1MTV; 12-NOV-97.  
DR PDB; 1MTW; 12-NOV-97.  
DR PDB; 1TPS; 26-JAN-95.

DR PDB; 1TYN; 26-JAN-95.  
DR PDB; 1TAM; 24-JUN-97.  
DR PDB; 1A07; 25-FEB-98.  
DR PDB; 1TIO; 23-SEP-98.  
DR PDB; 2TIO; 30-SEP-98.  
DR PDB; 1XUF; 16-DEC-98.  
DR PDB; 1XUG; 16-DEC-98.  
DR PDB; 1XUH; 11-NOV-98.  
DR PDB; 1XUI; 11-NOV-98.  
DR PDB; 1XUJ; 11-NOV-98.  
DR PDB; 1XUK; 11-NOV-98.  
DR PDB; 1AUJ; 14-OCT-98.  
DR PDB; 1AZ8; 13-JAN-99.  
DR PDB; 1BJV; 13-JAN-99.  
DR PDB; 2BZA; 23-MAR-99.  
DR PDB; 1CES; 23-MAR-99.  
DR MEROPS; S01.151; -;  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
DR KX Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen; Signal;  
KW 3D-structure.  
FT NON\_TER 1 1  
FT SIGNAL <1 14  
FT PROPEP 15 20 ACTIVATION PEPTIDE.  
FT CHAIN 21 243 TRYPSIN, CATIONIC.  
FT CHAIN 21 145 ALPHA-TRYPSIN CHAIN 1.  
FT CHAIN 146 243 ALPHA-TRYPSIN CHAIN 2.  
FT CA\_BIND 72 82  
FT BINDING 191 192 SUBSTRATE.  
FT BINDING 194 195 SUBSTRATE.  
FT BINDING 197 197 SUBSTRATE.  
FT DISULFID 27 157  
FT DISULFID 45 61  
FT DISULFID 129 230  
FT DISULFID 136 203  
FT DISULFID 168 182  
FT DISULFID 193 217  
FT ACT\_SITE 60 60  
FT ACT\_SITE 104 104 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 197 197 CHARGE RELAY SYSTEM.  
FT SITE 191 191 CHARGE RELAY SYSTEM.  
FT STRAND 22 22 REQUIRED FOR SPECIFICITY.  
FT STRAND 25 26  
FT TURN 29 30  
FT TURN 33 34  
FT STRAND 35 39  
FT STRAND 43 51  
FT TURN 52 53  
FT STRAND 54 57  
FT HELIX 59 61  
FT STRAND 67 70  
FT TURN 75 76  
FT STRAND 83 92  
FT TURN 94 95  
FT STRAND 97 97  
FT TURN 98 101  
FT STRAND 102 102  
FT TURN 103 103  
FT STRAND 106 110  
FT STRAND 124 124  
FT TURN 132 133  
FT STRAND 135 140  
FT STRAND 156 162  
FT HELIX 165 171  
FT TURN 173 174  
FT TURN 178 179  
FT STRAND 180 183  
FT TURN 186 187

Query Match 100.0%; Score 107; DB 1; Length 243;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20  
 |||||  
 Db 21 IVGGYTCGANTVPYQVSLNS 40

RESULT 2  
 TRYP\_PIG  
 ID TRYP\_PIG STANDARD; PRT; 231 AA.  
 AC P00761;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Trypsin precursor (EC 3.4.21.4).  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9622;  
 RN [1]  
 RP SEQUENCE OF 1-10.  
 RA Charles M., Rovey M., Guidoni A.A., Desmuelle P.;  
 RT "On trypsinogen and trypsin of pig.";  
 RL Biochim. Biophys. Acta 69:115-129(1963).  
 RN [2]  
 RP SEQUENCE OF 9-231.  
 RX MEDLINE=73258692; PubMed=4738933;  
 RA Hermodson M.A., Ericsson L.H., Neurath H., Walsh K.A.;  
 RT "Determination of the amino acid sequence of porcine trypsin by  
 RT sequenator analysis.";  
 RL Biochemistry 12:3146-3153(1973).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).  
 RX MEDLINE=93187998; PubMed=8445634;  
 RA Huang Q., Liu S., Tang Y.;  
 RT "Refined 1.6-A resolution crystal structure of the complex formed  
 RT between porcine beta-trypsin and MCTI-A, a trypsin inhibitor of the  
 RT squash family. Detailed comparison with bovine beta-trypsin and its  
 RT complex.";  
 RL J. Mol. Biol. 229:1022-1030(1993).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).  
 RX MEDLINE=92201369; PubMed=1551419;  
 RA Huang Q., Liu S., Tang Y., Zeng F., Qian R.;  
 RT "Amino acid sequencing of a trypsin inhibitor by refined 1.6 A X-ray  
 RT crystal structure of its complex with porcine beta-trypsin.";  
 RL FEBS Lett. 297:143-146(1992).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).  
 RX MEDLINE=95035057; PubMed=7947985;  
 RA Huang Q., Wang Z., Li Y., Liu S., Tang Y.;  
 RT "Refined 1.8-A resolution crystal structure of the porcine epsilon-on-  
 RT trypsin.";  
 RL Biochim. Biophys. Acta 1209:77-82(1994).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF COMPLEX WITH LDTI.  
 RX MEDLINE=97390427; PubMed=9242660;  
 RA Stubbs M.T., Morenweiser R., Sturzebecher J., Bauer M., Bode W.,  
 RA Huber R., Piechotka G.P., Matschner G., Sommerhoff C.P., Fritz H.,  
 RA Auerwald E.A.;  
 RT "The three-dimensional structure of recombinant leech-derived  
 RT trypsin inhibitor in complex with trypsin. Implications for the  
 RT structure of human mast cell tryptase and its inhibition.";  
 RL J. Biol. Chem. 272:19931-19937(1997).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (2.03 ANGSTROMS) OF COMPLEX WITH LDTI.  
 RX MEDLINE=98046095; PubMed=9384562;  
 RA di Marco S., Priestle J.P.;  
 RT "Structure of the complex of leech-derived trypsinase inhibitor (LDTI)  
 RT with trypsin and modelling of the LDTI-trypsinase system.";

RL Structure 5:1465-1474(1997).  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 DR PIR; A00947; TRPGTR.  
 DR PDB; 1MCT; 31-JAN-94.  
 DR PDB; 1AKS; 12-FEB-97.  
 DR PDB; 1BPT; 07-FEB-95.  
 DR PDB; 1TFX; 21-JAN-98.  
 DR PDB; 1LDT; 20-MAY-98.  
 DR PDB; 1ANI; 01-JUL-98.  
 DR PDB; 1AVW; 18-NOV-98.  
 DR PDB; 1AVX; 18-NOV-98.  
 DR MEROPS; S01.151; -.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRYD\_SPC; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen;  
 KW 3D-structure.  
 FT PROPEP 1 8  
 FT CHAIN 9 231  
 FT ACT\_SITE 48  
 FT ACT\_SITE 92 92  
 FT ACT\_SITE 185 185  
 FT DISULFID 15 145  
 FT DISULFID 33 49  
 FT DISULFID 117 218  
 FT DISULFID 124 191  
 FT DISULFID 156 170  
 FT DISULFID 181 205  
 FT SITE 179 179  
 FT VARIANT 20 20  
 FT STRAND 10 10  
 FT STRAND 13 14  
 FT TURN 17 18  
 FT TURN 21 22  
 FT STRAND 23 27  
 FT STRAND 31 39  
 FT TURN 40 41  
 FT STRAND 42 45  
 FT HELIX 47 49  
 FT STRAND 55 58  
 FT STRAND 62 62  
 FT TURN 63 64  
 FT STRAND 71 80  
 FT TURN 82 83  
 FT TURN 86 88  
 FT TURN 90 91  
 FT STRAND 94 98  
 FT STRAND 112 112  
 FT TURN 120 121  
 FT STRAND 123 128  
 FT STRAND 142 142  
 FT STRAND 144 150  
 FT HELIX 153 159  
 FT TURN 161 163  
 FT TURN 166 167  
 FT STRAND 168 171  
 FT TURN 174 175  
 FT STRAND 179 179  
 FT TURN 182 183  
 FT TURN 185 186  
 FT STRAND 188 191  
 FT TURN 192 193  
 FT STRAND 194 202  
 FT STRAND 207 207  
 FT TURN 208 209  
 FT STRAND 210 210

REQUIRED FOR SPECIFICITY (BY SIMILARITY).  
 I -> V.

```

FT STRAND      212 216
FT HELIX       217 219
FT HELIX       221 230
SQ SEQUENCE    231 AA; 24409 MW; A0A125CF7FC138C2 CRC64;

Query Match      89.7%; Score 96; DB 1; Length 231;
Best Local Similarity 85.0%; Pred. No. 7e-08;
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20
DB 9 IVGGYTCGAANSIPYQVSLNS 28

RESULT 3
TRY1 CANFA
ID TRY1 CANFA STANDARD; PRT; 246 AA.
AC P06871;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin, cationic precursor (EC 3.4.21.4).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86284628; PubMed=3841794;
RA Pinsky S.D., Laforge K.S., Scheele G.;
RT "Differential regulation of trypsinogen mRNA translation: full-length
RT mRNA sequences encoding two oppositely charged trypsinogen isoenzymes
RT in the dog pancreas."
RL Mol. Cell. Biol. 5:2669-2676(1985).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M11530; AAA30900.1; -.
DR PIR; B26273; TRDGC.
DR HSP; P00761; LEPT.
DR MEROPS; S01.151; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser.protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRY_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
KW Multigene family.
FT SIGNAL      1 15
FT PROPEP     16 23 ACTIVATION PEPTIDE.
FT CHAIN      24 246 TRYPSIN, CATIONIC.
FT ACT_SITE   63 63 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE  107 107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE  200 200 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID   30 160 BY SIMILARITY.
FT DISULFID   48 64 BY SIMILARITY.
FT DISULFID  132 233 BY SIMILARITY.
FT DISULFID  139 206 BY SIMILARITY.
FT DISULFID  171 185 BY SIMILARITY.
FT DISULFID  186 220 BY SIMILARITY.
FT SITE       194 194 REQUIRED FOR SPECIFICITY (BY SIMILARITY).

SQ SEQUENCE    246 AA; 26170 MW; E9E5A1DE2391B8BB CRC64;

Query Match      86.0%; Score 92; DB 1; Length 246;
Best Local Similarity 85.0%; Pred. No. 3.2e-07;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20
DB 24 IVGGYTCRSNSVPYQVSLNS 43

RESULT 4
TRY2 RAT
ID TRY2 RAT STANDARD; PRT; 246 AA.
AC P00763;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin II, anionic precursor (EC 3.4.21.4) (Pretypsinogen II).
GN TRY2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85054880; PubMed=6094547;
RA Craik C.S., Choo Q.L., Swift G.H., Quinto C., McDonald R.J.,
RA Rutter W.J.;
RT "Structure of two related rat pancreatic trypsin genes."
RL J. Biol. Chem. 259:14255-14264(1984).
RN [2]
RP SEQUENCE OF 9-246 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Pancreas;
RX MEDLINE=82265624; PubMed=6896710;
RA McDonald R.J., Stary S.J., Swift G.H.;
RT "Two similar but nonallelic rat pancreatic trypsinogens. Nucleotide
RT sequences of the cloned cDNAs."
RL J. Biol. Chem. 257:9724-9732(1982).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.59 ANGSTROMS).
RX MEDLINE=91351998; PubMed=1881877;
RA Earnest T., Fauman E., Craik C.S., Stroud R.;
RT "1.59-A structure of trypsin at 120 K: comparison of low temperature
RT and room temperature structures."
RL Proteins 10:171-187(1991).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=96214506; PubMed=8634241;
RA Brinen L.S., Willett W.S., Craik C.S., Fletterick R.J.;
RT "X-ray structures of a designed binding site in trypsin show metal-
RT dependent geometry."
RL Biochemistry 35:5999-6009(1996).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V01274; CAA24581.1; -.
DR EMBL; L00131; AAA98517.1; -.
DR EMBL; L00130; AAA98517.1; JOINED.
DR PIR; A22657; TRRT2.
DR PDB; 1ANB; 01-APR-97.
DR PDB; 1ANC; 01-APR-97.
DR PDB; 1AND; 01-APR-97.
DR PDB; 1ANE; 01-APR-97.

```

```

DR PDB: 1AMH; 24-DEC-97.
DR PDB: 1SLU; 11-JUL-96.
DR PDB: 1SLV; 11-JUL-96.
DR PDB: 1SLW; 11-JUL-96.
DR PDB: 1SLX; 11-JUL-96.
DR PDB: 1DFO; 07-JUL-97.
DR PDB: 3TGI; 23-DEC-98.
DR PDB: 3TGT; 23-DEC-98.
DR MEROPS; S01.258; .
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser. protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
KW Multigene family; 3D-structure.
FT SIGNAL 1 15
FT PROPEP 16 23 ACTIVATION PEPTIDE.
FT CHAIN 24 246 TRYPSIN II, ANIONIC.
FT ACT_SITE 63 63 CHARGE RELAY SYSTEM.
FT ACT_SITE 107 107 CHARGE RELAY SYSTEM.
FT ACT_SITE 200 200 CHARGE RELAY SYSTEM.
FT DISULFID 30 160 CHARGE RELAY SYSTEM.
FT DISULFID 48 64 BY SIMILARITY.
FT DISULFID 132 233 BY SIMILARITY.
FT DISULFID 139 206 BY SIMILARITY.
FT DISULFID 171 185 BY SIMILARITY.
FT DISULFID 196 220 BY SIMILARITY.
FT SITE 194 194 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
FT CONFLICT 84 84 N -> D (IN REF. 1).
FT CONFLICT 88 88 V -> I (IN REF. 1).
SQ SEQUENCE 246 AA; 26228 MW; A8D3630809AEB606 CRC64;

Query Match 84.1%; Score 90; DB 1; Length 246;
Best Local Similarity 85.0%; Pred. No. 6.5e-07;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20
Db 24 IVGGYTCGANTVPYQVSLNS 43
||||||| :|||||||

RESULT 5
TRY2_BOVIN
ID TRY2_BOVIN STANDARD; PRT; 247 AA.
AC Q29463;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin, anionic precursor (EC 3.4.21.4).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Holstein-Friesian; TISSUE=Pancreas;
RX MEDLINE=91065383; PubMed=1701147;
RA le Huotou I., Wicker C., Guilloteau P., Toullec R., Puigserver A.;
RT "Isolation and nucleotide sequence of cDNA clone for bovine
pancreatic anionic trypsinogen. Structural identity within the
trypsin family.";
RL Eur. J. Biochem. 193:767-773(1990).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
-----
DR EMBL; X54703; CAA38513.1; .
DR HSP; P00763; LDPO.
DR MEROPS; S01.258; .
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser. protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT PROPEP 16 23 ACTIVATION PEPTIDE.
FT CHAIN 24 247 TRYPSIN, ANIONIC.
FT ACT_SITE 63 63 CHARGE RELAY SYSTEM.
FT ACT_SITE 107 107 CHARGE RELAY SYSTEM.
FT ACT_SITE 200 200 CHARGE RELAY SYSTEM.
FT DISULFID 30 160 BY SIMILARITY.
FT DISULFID 48 64 BY SIMILARITY.
FT DISULFID 132 233 BY SIMILARITY.
FT DISULFID 139 206 BY SIMILARITY.
FT DISULFID 171 185 BY SIMILARITY.
FT DISULFID 196 220 BY SIMILARITY.
FT SITE 194 194 REQUIRED FOR SPECIFICITY.
SQ SEQUENCE 247 AA; 26289 MW; 50A070495A7731DB CRC64;

Query Match 83.2%; Score 89; DB 1; Length 247;
Best Local Similarity 80.0%; Pred. No. 9.3e-07;
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20
Db 24 IVGGYTCGANTVPYQVSLNS 43
||||||| :|||||||

RESULT 6
TRY2_CANPA
ID TRY2_CANPA STANDARD; PRT; 247 AA.
AC P06872;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin, anionic precursor (EC 3.4.21.4).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86284628; PubMed=3841794;
RA Pinsky S.D., Laforge K.S., Scheele G.;
RT "Differential regulation of trypsinogen mRNA translation: full-length
mRNA sequences encoding two oppositely charged trypsinogen isoenzymes
in the dog pancreas.";
RL Mol. Cell. Biol. 5:2669-2676(1985).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)

```

```
CC EMBL; M11589; AAA30899.1; -.
DR PIR; A26273; TRDG
DR HSP; P00761; LDPO.
DR MEROPS; S01.174; -.
DR InterPro; IPR001314; Chymotrypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS0240; TRYP SIN DOM; 1.
DR PROSITE; PS00134; TRYP SIN HIS; 1.
DR PROSITE; PS00135; TRYP SIN SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
KW Multigene family.
FT SIGNAL 1 15 ACTIVATION PEPTIDE.
FT PROPEP 16 23 TRYP SIN, ANIONIC.
FT CHAIN 24 247
FT ACT_SITE 63 63 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 107 107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 200 200 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 30 160 BY SIMILARITY.
FT DISULFID 48 64 BY SIMILARITY.
FT DISULFID 132 233 BY SIMILARITY.
FT DISULFID 139 206 BY SIMILARITY.
FT DISULFID 171 185 BY SIMILARITY.
FT DISULFID 196 220 BY SIMILARITY.
FT SITE 194 194 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 247 AA; 28423 MW; 374E9D31D6DB8EAF CRC64;

Query Match 81.3%; Score 87; DB 1; Length 247;
Best Local Similarity 80.0%; Pred. No. 1.9e-06;
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20
Db 24 IVGGYTCGANTVPYQVSLNS 43
||||| :|||||

RESULT 7
TRY3_HUMAN STANDARD; PRT; 247 AA.
AC P15951;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin III precursor (EC 3.4.21.4).
GN PRSS3 OR TRY3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=50221895; PubMed=2326201;
RA Tani T., Kawashima I., Mita K., Takiguchi Y.;
RL Nucleic Acids Res. 18:1631-1631(1990).
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X15505; CAA33527.1; -.
DR PIR; S12764; S12764.
```

```
DR HSP; P00761; 1EPT.
DR MEROPS; S01.174; -.
DR Genew; HGNC:9486; PRSS3.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser.protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS0240; TRYP SIN DOM; 1.
DR PROSITE; PS00134; TRYP SIN HIS; 1.
DR PROSITE; PS00135; TRYP SIN SER; FALSE_NEG
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
KW Multigene family.
FT SIGNAL 1 15 ACTIVATION PEPTIDE.
FT PROPEP 16 23 TRYP SIN III.
FT CHAIN 24 247
FT ACT_SITE 63 63 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 107 107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 200 200 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 30 160 BY SIMILARITY.
FT DISULFID 48 64 BY SIMILARITY.
FT DISULFID 139 206 BY SIMILARITY.
FT DISULFID 171 185 BY SIMILARITY.
FT SITE 194 194 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 247 AA; 26776 MW; 697DE163FICE80D6 CRC64;

Query Match 81.3%; Score 87; DB 1; Length 247;
Best Local Similarity 80.0%; Pred. No. 1.9e-06;
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20
Db 24 IVGGYTCGANTVPYQVSLNS 43
||||| :|||||

RESULT 8
TRY4_HUMAN STANDARD; PRT; 304 AA.
AC P35030; Q15665; Q9UQV3;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin IV precursor (EC 3.4.21.4) (Brain trypsinogen)
DE (Mesotrypsinogen).
GN PRSS4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A AND B).
RC TISSUE=Brain;
RX MEDLINE=94123994; PubMed=8294000;
RA Wiegand U., Corbach S., Minn A., Kang J., Mueller-Hill B.;
RT "Cloning of the cDNA encoding human brain trypsinogen and
RT characterization of its product.";
RL Gene 136:167-175(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM C).
RA Fukuda S.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- ALTERNATIVE PRODUCTS: 3 isoforms; A (shown here), B and C; seem
CC to be produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: PANCREAS AND BRAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X15505; CAA33527.1; -.
DR PIR; S12764; S12764.
```

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC CC EMBL; X73781; CAB58178.1; -
CC CC EMBL; X71345; CAA50484.1; -
CC CC EMBL; D45417; BAA08257.1; -
CC CC PIR; S13496; S33496.
CC CC HSSP; P07477; ITRN.
CC CC DR INTERPRO; IPR001314; Chymotrypsin.
CC CC DR InterPro; IPR001254; Ser. protease_Try.
CC CC DR Pfam; PF00089; trypsin; 1.
CC CC DR PRINTS; PR00722; CHYMOTRYPSIN.
CC CC DR PROSITE; PS00240; TRYPSIN_DOM; 1.
CC CC DR PROSITE; PS00134; TRYPSIN_HIS; 1.
CC CC DR PROSITE; PS00135; TRYPSIN_SER; FALSE NEG.
CC CC KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
CC CC Multigene family; Alternative splicing.
CC CC FT SIGNAL 1 ?
CC CC FT PROPEP 2 80
CC CC FT CHAIN 81 304
CC CC FT ACT_SITE 120 120
CC CC FT ACT_SITE 164 164
CC CC FT ACT_SITE 257 257
CC CC FT DISULFID 87 217
CC CC FT DISULFID 105 121
CC CC FT DISULFID 196 263
CC CC FT DISULFID 228 242
CC CC FT DISULFID 253 277
CC CC FT SITE 251 251
CC CC FT VARSPIC 1 45
CC CC FT VARSPIC 1 70
CC CC FT CONFLICT 89 89
CC CC FT SEQUENCE 304 AA; 32499 MW; 4C4316C3F1D0FFC CRC64;

Query Match 81.3%; Score 87; DB 1; Length 304;
Best Local Similarity 80.0%; Pred. No. 2.3e-06;
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPQVSLS 20
Db 81 IVGGYTCENSIPQVSLS 100
||||||| :|||
||||||| :|||

RESULT 9
TRY1 RAT STANDARD; PRT; 246 AA.
AC P00762;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin 1, anionic precursor (EC 3.4.21.4) (Pretrypsinogen 1).
GN TRY1.
OS Rattus norvegicus (Rat).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
CC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=82265624; PubMed=696710;
RA McDonald R.J., Stary S.J., Swift G.H.;
RT "Two similar but nonallelic rat pancreatic trypsinogens. Nucleotide
RL sequences of the cloned cDNAs."
RJ J. Biol. Chem. 257:9724-9732(1982).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85054880; PubMed=6094547;
RA Craik C.S., Choo Q.L., Swift G.H., Quinto C., McDonald R.J.,

```

```

RA Rutter W.J.;
RT "Structure of two related rat pancreatic trypsin genes.";
RL J. Biol. Chem. 259:14255-14264(1984).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE=87292123; PubMed=3112942;
RA Sprang S., Standing T., Fletterick R.J., Stroud R.M., Finer-Moore J.,
RA Xuong N.-H., Hamlin R., Rutter W.J., Craik C.S.;
RT "The three-dimensional structure of Asn102 mutant of trypsin: role of
RT Asn102 in serine protease catalysis.";
RL Science 237:905-909(1987).
CC CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC CC -1- SUBCELLULAR LOCATION: Extracellular.
CC CC -1- MISCELLANEOUS: THIS SEQUENCE REPRESENTS THE PRECURSOR OF THE MAJOR
CC CC FORM OF TRYPSIN PRODUCED BY THE ADULT PANCREAS.
CC CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).

```

```

EMBL; V01273; CAA24580.1; -
EMBL; J00778; AAA98518.1; -
PIR; A00948; TRRT1.
PDB; 1TRM; 15-JUL-93.
PDB; 2TRM; 16-JUL-88.
PDB; 1BBA; 30-APR-94.
PDB; 1BRB; 31-JUL-94.
PDB; 1BRC; 31-MAY-94.
MEROPS; S01.094; -.
InterPro; IPR001314; Chymotrypsin.
InterPro; IPR001254; Ser. protease_Try.
Pfam; PF00089; trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
SMART; SM00020; TRYD_SPC; 1.
PROSITE; PS00240; TRYPSIN_DOM; 1.
PROSITE; PS00134; TRYPSIN_HIS; 1.
PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
KW Multigene family; 3D-structure.
FT SIGNAL 1 15
FT PROPEP 16 23
FT CHAIN 24 246
FT ACT_SITE 63 63
FT ACT_SITE 107 107
FT ACT_SITE 200 200
FT DISULFID 30 160
FT DISULFID 48 64
FT DISULFID 132 233
FT DISULFID 139 206
FT DISULFID 171 185
FT DISULFID 196 220
FT SITE 194 194
FT STRAND 25 25
FT STRAND 28 29
FT TURN 32 33
FT TURN 36 37
FT STRAND 38 42
FT STRAND 46 54
FT TURN 55 56
FT STRAND 57 60
FT HELIX 62 64
FT STRAND 70 73
FT TURN 77 77
FT TURN 78 79
FT STRAND 86 95
FT TURN 97 98
FT STRAND 100 100
FT TURN 101 104

```

REQUIRED FOR SPECIFICITY.





CC EMBL; M26612; AAA61231.1; -;  
DR EMBL; U70137; AAC50728.1; -;  
DR EMBL; AF114534; AAG30943.1; -;  
DR EMBL; AF115309; AAG30947.1; -;  
DR EMBL; AF115310; AAG30948.1; -;  
DR EMBL; AF115311; AAG30949.1; -;  
DR PIR; A25852; A25852.  
DR PDB; 1TRN; 03-JUN-95.  
DR PDB; 1FXV; 17-JUN-98.  
DR MEROPS; S01.151; -;  
DR Genew; HGNC:9475; PRS51.  
DR MIM; 276000; -;  
DR MIM; 167800; -;  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRYD\_SPC; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
DR Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;  
KW Multigene family; Phosphorylation; 3D-structure; Disease mutation.  
FT SIGNAL 1 15  
FT PROPEP 16 23 ACTIVATION PEPTIDE.  
FT CHAIN 24 247 TRYPSIN I.  
FT ACT\_SITE 63 63 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 107 107 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 200 200 CHARGE RELAY SYSTEM.  
FT DISULFID 30 160  
FT DISULFID 48 64  
FT DISULFID 139 206  
FT DISULFID 171 185  
FT DISULFID 196 220  
FT MOD\_RES 154 154  
FT SITE 194 194  
FT VARIANT 16 16  
FT VARIANT 22 22  
FT VARIANT 23 23  
FT VARIANT 29 29  
FT VARIANT 29 29  
FT VARIANT 104 104  
FT VARIANT 116 116  
FT VARIANT 122 122  
FT VARIANT 122 122  
FT VARIANT 122 122  
FT VARIANT 122 122  
FT VARIANT 139 139  
FT VARIANT 139 139  
FT SEQUENCE 247 AA; 26558 MW; DD49A487B8062813 CRC64;  
Query Match 79.4%; Score 85; DB 1; Length 247;  
Best Local Similarity 80.0%; Pred. No. 3.9e-06;  
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20  
DB 24 IVGGYNCBNSVPYQVSLNS 43  
RESULT 11  
TRY3 CHICK STANDARD; PRT; 248 AA.  
AC O90629;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Trypsin II-p29 precursor (EC 3.4.21.4).  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RX MEDLINE=95251611; PubMed=7733885;  
RA Wang K., Gan L., Lee L., Hood L.E.;  
RT "Isolation and characterization of the chicken trypsinogen gene family";  
RL Biochem J. 307:471-479(1995).  
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
CC -1- SUBCELLULAR LOCATION: Extracellular.  
CC -1- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE PANCREAS WHILE LOWER LEVELS ARE FOUND IN THE LIVER, SPLEEN AND THYMUS.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sit.ch/announce/ or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; U15157; AAA79914.1; -;  
DR HSPP; P00763; LDPO.  
DR MEROPS; S01.151; -;  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRYD\_SPC; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;  
KW Multigene family.  
FT SIGNAL 1 16 BY SIMILARITY.  
FT PROPEP 17 25 ACTIVATION PEPTIDE (BY SIMILARITY).  
FT CHAIN 26 248 TRYPSIN II-P29.  
FT ACT\_SITE 65 65 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 32 162 BY SIMILARITY.  
FT DISULFID 50 66 BY SIMILARITY.  
FT DISULFID 134 235 BY SIMILARITY.  
FT DISULFID 141 208 BY SIMILARITY.  
FT DISULFID 173 187 BY SIMILARITY.  
FT DISULFID 198 222 BY SIMILARITY.  
FT SITE 196 196 REQUIRED FOR SPECIFICITY (BY SIMILARITY).  
SQ SEQUENCE 248 AA; 26622 MW; E5E16B07622B588E CRC64;  
Query Match 79.4%; Score 85; DB 1; Length 248;  
Best Local Similarity 80.0%; Pred. No. 3.9e-06;  
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20

```
Db 26 IVGGYTCPEHSVPYQVSLNS 45
||||| :|||||
RESULT 12
TRY2_XENLA STANDARD; PRT; 244 AA.
AC P70059;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin precursor (EC 3.4.21.4)
OS Xenopus laevis (African clawed frog)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RN SEQUENCE FROM N.A.
RP Wang K., Lytle L., Gan L., Hood L.E.;
RA Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RL -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg|-Xaa, Lys|-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; U72330; AAB17274.1; -
DR HSSP; P00763; LDPO.
DR MEROPS; S01.258; -
DR InterPro; IPR001314; Chymotrypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Digestion; Zymogen; Signal;
KW Multigene family.
FT SIGNAL 1 15 BY SIMILARITY.
FT PROPEP 16 21 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 22 244 TRYPsin.
FT ACT_SITE 61 61 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 105 105 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 198 198 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 28 158 BY SIMILARITY.
FT DISULFID 46 62 BY SIMILARITY.
FT DISULFID 130 204 BY SIMILARITY.
FT DISULFID 137 204 BY SIMILARITY.
FT DISULFID 169 183 BY SIMILARITY.
FT DISULFID 194 218 BY SIMILARITY.
FT SITE 192 192 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 244 AA; 26079 MW; C63F29CB3300B323 CRC64;
Query Match 78.5%; Score 84; DB 1; Length 244;
Best Local Similarity 75.0%; Pred. No. 5.5e-06;
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Qy 1 IVGGYTCGANTVPYQVSLNS 20
||||| :|||||
Db 22 IVGGFTCAKNAVYPYQVSLNA 41
RESULT 13
TRY2_HUMAN STANDARD; PRT; 247 AA.
AC P70478;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin II precursor (EC 3.4.21.4) (Anionic trypsinogen).
GN PRSS2 OR TRY2 OR TRY2P.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=86221712; PubMed=3011602;
RA Emi M., Nakamura Y., Ogawa M., Yamamoto T., Nishide T., Mori T.,
RA Matubara K.;
RT "Cloning, characterization and nucleotide sequences of two cDNAs
RT encoding human pancreatic trypsinogens.";
RL Gene 41:305-310(1986).
RN [2]
RN SEQUENCE OF 16-49.
RP MEDLINE=90091010; PubMed=2598466;
RA Kimland M., Russick C., Marks W.H., Borgstroem A.;
RT "Immunoreactive anionic and cationic trypsin in human serum.";
RL Clin. Chim. Acta 184:31-46(1989).
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg|-Xaa, Lys|-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; M27602; AAA61232.1; -
DR FIR; B25852; B25852.
DR HSSP; P00763; LDPO.
DR MEROPS; S01.258; -
DR Genew; HGNC:9483; PRSS2.
DR MIM; 601564; -
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
KW Multigene family.
FT SIGNAL 1 15 ACTIVATION PEPTIDE.
FT PROPEP 16 23 TRYPsin II.
FT CHAIN 24 247 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 63 63 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 107 107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 200 200 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 30 160 BY SIMILARITY.
FT DISULFID 48 64 BY SIMILARITY.
FT DISULFID 171 185 BY SIMILARITY.
FT DISULFID 196 220 BY SIMILARITY.
FT SITE 194 194 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 247 AA; 26488 MW; 82B0F41EB9E3D5DB CRC64;
Query Match 78.5%; Score 84; DB 1; Length 247;
Best Local Similarity 80.0%; Pred. No. 5.6e-06;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Qy 1 IVGGYTCGANTVPYQVSLNS 20
||||| :|||||
Db 24 IVGGYICEENSVPYQVSLNS 43
```

```

RESULT 14
TRY3 RAT
ID TR3 RAT STANDARD; PRT; 247 AA.
AC P08426;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin III, cationic precursor (EC 3.4.21.4) (Pretrypsinogen III).
GN TRV3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=87271609; PubMed=3607011;
RX Fletcher T.S., Alhadeff M., Craik C.S., Lutzman C.;
RT "Isolation and characterization of a cDNA encoding rat cationic
RT trypsinogen.";
RL Biochemistry 26:3081-3086(1987).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M16624; AAA1985.1; .
CC PIR; A27547; A27547.
CC HSP; P00763; LDPO.
CC MEROPS; S01.056; .
CC InterPro; IPR001314; Chymotrypsin.
CC InterPro; IPR001254; Ser protease_Try.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; TRYD_SPC; 1.
CC PROSITE; PS50240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
KW Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 24 ACTIVATION PEPTIDE.
FT CHAIN 16 24 TRYPSIN III, CATIONIC.
FT ACT_SITE 25 247 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 64 64 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 108 108 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 201 201 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 31 161 BY SIMILARITY.
FT DISULFID 49 65 BY SIMILARITY.
FT DISULFID 133 234 BY SIMILARITY.
FT DISULFID 140 207 BY SIMILARITY.
FT DISULFID 172 186 BY SIMILARITY.
FT DISULFID 197 221 BY SIMILARITY.
FT SITE 195 195 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 247 AA; 26269 MW; D74892BAA584E4A8 CRC64;

Query Match 78.5%; Score 84; DB 1; Length 247;
Best Local Similarity 75.0%; Pred. No. 5.6e-06;
Matches 15; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20
Db 25 IVGGYTCQNSLPYQVSLNA 44

RESULT 15
TRY2 MOUSE
ID TRY2 MOUSE STANDARD; PRT; 246 AA.
AC P07146;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin II, anionic precursor (EC 3.4.21.4) (Pretrypsinogen II).
GN TRY2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RX STRAIN=A/J;
RX MEDLINE=87066713; PubMed=3641189;
RX Stevenson B.J., Hagenbueschle O., Wellauer P.K.;
RT "Sequence organisation and transcriptional regulation of the mouse
RT elastase II and trypsin genes.";
RL Nucleic Acids Res. 14:8307-8330(1986).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X04574; CAA28243.1; .
CC EMBL; X04577; CAA28245.1; .
CC PIR; B55528; B25528.
CC HSP; P00763; LDPO.
CC MEROPS; S01.064; .
CC MGI; MGI:102759; Try2.
CC InterPro; IPR001314; Chymotrypsin.
CC InterPro; IPR001254; Ser protease_Try.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; TRYD_SPC; 1.
CC PROSITE; PS50240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal.
KW SIGNAL 1 15
FT PROPEP 16 23 ACTIVATION PEPTIDE.
FT CHAIN 16 23 TRYPSIN II, ANIONIC.
FT ACT_SITE 24 246 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 63 63 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 107 107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 200 200 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 30 160 BY SIMILARITY.
FT DISULFID 48 64 BY SIMILARITY.
FT DISULFID 132 233 BY SIMILARITY.
FT DISULFID 139 206 BY SIMILARITY.
FT DISULFID 171 185 BY SIMILARITY.
FT DISULFID 196 220 BY SIMILARITY.
FT SITE 194 194 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 246 AA; 26203 MW; CEF8C97AAC2D07AD CRC64;

Query Match 76.6%; Score 82; DB 1; Length 246;
Best Local Similarity 75.0%; Pred. No. 1.1e-05;
Matches 15; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20
Db 24 IVGGYTCRESSVPYQVSLNA 43

Search completed: February 12, 2003, 10:23:12
Job time : 2.68657 secs

```

**THIS PAGE BLANK (USPTO)**



```

DE GN Trypsinogen 16.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA Rowen L., Hood L.;
RT "Comparison between strains Balb/C and 129 in a region of the mouse T
RL cell receptor beta locus.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=21103195; PubMed=11160223;
RA Chen F., Rowen L., Hood L., Rothenberg E.V.;
RT "Differential transcriptional regulation of individual TCR Vbeta
RL segments before gene rearrangement.";
RL J. Immunol. 166:1771-1780(2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AF107342; AAC79093.1; -.
DR EMBL; AE000685; AAB69088.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.063; -.
DR MGD; MGI:2148749; Tcvg16.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRY SP; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KM Hydrolase; Serine protease.
SQ SEQUENCE 246 AA; 26134 MW; 34B173B18CA2F463 CRC64;

Query Match 84.1%; Score 90; DB 11; Length 246;
Best Local Similarity 85.0%; Pred. No. 2.7e-06;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IVGGYTCGANTVPYQVSLNS 20
Db 24 IVGGYTCRENSVPYQVSLNS 43

RESULT 3
Q9ROT7 PRELIMINARY; PRT; 246 AA.
AC Q9ROT7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pancreatic trypsin (0910001B19RIK protein) (Trypsinogen 8).
GN TD 0910001B19RIK OR TRYPSINOGEN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129SVJ;
RX MEDLINE=99436155; PubMed=10506205;
RA Ohmura K., Kohno N., Kobayashi Y., Yamagata K., Sato S.,
RA Kashiwabara S., Baba T.;
RT "A homologue of pancreatic trypsin is localized in the acrosome of
RT mammalian sperm and is released during acrosome reaction.";
RL J. Biol. Chem. 274:29426-29432(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=STOMACH, AND SPLEEN;

```

```

RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Komoto H., Adachi J., Fukuda S.,
RA Azawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Giesi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pasole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Maehima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayaishizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=21103195; PubMed=11160223;
RA Chen F., Rowen L., Hood L., Rothenberg E.V.;
RT "Differential transcriptional regulation of individual TCR Vbeta
RT segments before gene rearrangement.";
RL J. Immunol. 166:1771-1780(2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB017032; BAA74761.1; -.
DR EMBL; AK008667; BAB25821.1; -.
DR EMBL; AK003064; BAB2542.1; -.
DR EMBL; AB000664; AAB69056.1; -.
DR HSSP; P00763; ISLU.
DR MEROPS; S01.057; -.
DR MGD; MGI:1913350; 0910001B19RIK.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRY SP; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KM Hydrolase; Serine protease.
SQ SEQUENCE 246 AA; 26274 MW; B6A9F4C99079633F CRC64;

Query Match 84.1%; Score 90; DB 11; Length 246;
Best Local Similarity 85.0%; Pred. No. 2.7e-06;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IVGGYTCGANTVPYQVSLNS 20
Db 24 IVGGYTCRENSVPYQVSLNS 43

RESULT 4
Q9QUK9 PRELIMINARY; PRT; 246 AA.
AC Q9QUK9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2002 (TrEMBLrel. 20, Last annotation update)
DE TESP4 (0910001B19RIK protein) (Trypsinogen 9).
GN TC 0910001B19RIK OR TRYPSINOGEN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=TESTIS;
RX MEDLINE=99436155; PubMed=10506205;
RA Ohmura K., Kohno N., Kobayashi Y., Yamagata K., Sato S.,
RA Kaishiwabara S., Baba T.;
RT "A homologue of pancreatic trypsin is localized in the acrosome of
RT mammalian sperm and is released during acrosome reaction.";
RL J. Biol. Chem. 274:29426-29432(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schrim L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=21103195; PubMed=11160223;
RA Chen F., Rowen L., Hood L., Rothenberg E.V.;
RT "Differential transcriptional regulation of individual TCR Vbeta
RT segments before gene rearrangement.";
RL J. Immunol. 166:1771-1780(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB009661; BAA85187.1; -
DR EMBL; AB017031; BAA74760.1; -
DR EMBL; AK007843; BAB25300.1; -
DR EMBL; AE000664; AAB69057.1; -
DR HSSP; P00763; IDPO.
DR MEROPS; S01.058; -
DR MGD; MGI:1913350; 0910001B19Rik.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 246 AA; 26277 MW; 915C92353EC42809 CRC64;

Query Match 83.2%; Score 89; DB 11; Length 246;
Best Local Similarity 80.0%; Pred. No. 3.9e-06;
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20
Db 24 IVGGYTCRENSIPYQVSLNS 43
||||| |:::|||||
1 IVGGYTCGANTVPYQVSLNS 20
24 IVGGYTCRENSIPYQVSLNS 43

RESULT 5
Q9CFN9
ID Q9CFN9 PRELIMINARY; PRT; 247 AA.
AC Q9CFN9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)

TISSUE=TESTIS;
RX MEDLINE=99436155; PubMed=10506205;
RA Ohmura K., Kohno N., Kobayashi Y., Yamagata K., Sato S.,
RA Kaishiwabara S., Baba T.;
RT "A homologue of pancreatic trypsin is localized in the acrosome of
RT mammalian sperm and is released during acrosome reaction.";
RL J. Biol. Chem. 274:29426-29432(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schrim L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=21103195; PubMed=11160223;
RA Chen F., Rowen L., Hood L., Rothenberg E.V.;
RT "Differential transcriptional regulation of individual TCR Vbeta
RT segments before gene rearrangement.";
RL J. Immunol. 166:1771-1780(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB009661; BAA85187.1; -
DR EMBL; AB017031; BAA74760.1; -
DR EMBL; AK007843; BAB25300.1; -
DR EMBL; AE000664; AAB69057.1; -
DR HSSP; P00763; IDPO.
DR MEROPS; S01.058; -
DR MGD; MGI:1913350; 0910001B19Rik.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 246 AA; 26277 MW; 915C92353EC42809 CRC64;

Query Match 83.2%; Score 89; DB 11; Length 246;
Best Local Similarity 80.0%; Pred. No. 3.9e-06;
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20
Db 24 IVGGYTCRENSIPYQVSLNS 43
||||| |:::|||||
1 IVGGYTCGANTVPYQVSLNS 20
24 IVGGYTCRENSIPYQVSLNS 43

RESULT 6
Q9D7Y7
ID Q9D7Y7 PRELIMINARY; PRT; 247 AA.
AC Q9D7Y7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)

```

```

DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE 2210010C04RIK
GN 2210010C04RIK
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=STOMACH;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR ENBL; AK008695; BAB25837.1; -.
DR HSP; P00763; IDPO.
DR MEROPS; S01.151; -.
DR MGD; MGI:1914623; 2210010C04RIK.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS0240; TRYPsin DOM; 1.
DR PROSITE; PS00134; TRYPsin HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPsin SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 247 AA; 26407 MW; 84D474DB322B3A55 CRC64;

Query Match 76.6%; Score 82; DB 11; Length 247;
Best Local Similarity 78.9%; Pred. No. 5e-05;
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 VGGYTCGANTVPYQVSLNS 20
Db ||||| :|||||||
26 VGGYTCORNALPYQVSLNS 44

RESULT 7
Q9CPN7 PRELIMINARY; PRT; 247 AA.
AC Q9CPN7
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2002 (TReMBLrel. 17, Last sequence update)
DE 1810009J06RIK protein (Trypsinogen 4).
GN 1810009J06RIK OR TRYPSINOGEN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR ENBL; AK007406; BAB25018.1; -.
DR EMBL; AE000663; AAB69055.1; -.
DR HSP; P00763; IDPO.
DR MEROPS; S01.129; -.
DR MGD; MGI:1920876; 1810009J06RIK.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS0240; TRYPsin DOM; 1.
DR PROSITE; PS00134; TRYPsin HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPsin SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 247 AA; 26503 MW; EDF5F7696833C7BC CRC64;

Query Match 75.7%; Score 81; DB 11; Length 247;
Best Local Similarity 78.9%; Pred. No. 7.2e-05;
Matches 15; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLN 19
Db ||||| :|||||||
24 IVGGYTCPKHSPYQVSLN 42

RESULT 8
Q91515 PRELIMINARY; PRT; 237 AA.
AC Q91515
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Trypsinogen (Fragment).
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.

```



```

RA Wang K., Gan L., Lee I., Roach J., Hood L.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; U25747; AAA75001.1; -.
DR HSSP; P35031; 1BIT.
DR MEROPS; S01.125; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease.
FT NON_TER 1
SQ SEQUENCE 237 AA; 25726 MW; 30D2DBAAC39080C2 CRC64;

Query Match 71.0%; Score 76; DB 13; Length 237;
Best Local Similarity 75.0%; Pred. No. 0.00042;
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLN 20
Db 16 IVGGYECRKNVAYQVSLN 35

RESULT 9
O42159 PRELIMINARY; PRT; 244 AA.
AC O42159;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsinogen B1 precursor (Fragment).
GN TrypB1.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
KW Hydrolase; Serine protease; Signal.
FT SIGNAL <1 13 POTENTIAL.
FT CHAIN 14 244 TRYPSIN B1.
FT SIGNAL <1 13 POTENTIAL.
FT CHAIN 14 245 TRYPSIN B2.
SQ SEQUENCE 244 AA; 25903 MW; C4582EE07E3B8007 CRC64;

Query Match 66.4%; Score 71; DB 13; Length 244;
Best Local Similarity 68.4%; Pred. No. 0.0027;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLN 19
Db 21 IVGGYECRKNVAYQVSLN 39

RESULT 10
O42160 PRELIMINARY; PRT; 245 AA.
AC O42160;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsinogen B2 precursor (Fragment).
GN TrypB2.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
KW Hydrolase; Serine protease; Signal.
FT NON_TER 1
SQ SEQUENCE FROM N.A.
RA Roach J.C.;
RT "The Molecular Evolution of the Vertebrate Trypsinogens.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AF011901; AAB69656.1; -.
DR HSSP; P00763; 1DPO.
DR MEROPS; S01.128; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease; Signal.
FT NON_TER 1
FT SIGNAL <1 13 POTENTIAL.
FT CHAIN 14 245 TRYPSIN B2.
FT SIGNAL <1 13 POTENTIAL.
FT CHAIN 14 245 TRYPSIN B2.
SQ SEQUENCE 245 AA; 26001 MW; 9A932508B96C93E CRC64;

Query Match 66.4%; Score 71; DB 13; Length 245;
Best Local Similarity 68.4%; Pred. No. 0.0027;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLN 19
Db 22 IVGGYECRKNVAYQVSLN 40

RESULT 11
O98TH0 PRELIMINARY; PRT; 240 AA.
AC O98TH0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsinogen.
OS Engraulis japonicus (Japanese anchovy).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Clupeomorpha; Engraulidae;
OC Engraulis.
OX NCBI_TaxID=42892;
RN [1]
RN [1]
SQ SEQUENCE FROM N.A.
RC TISSUE=PYLORIC CAECA;
RA Watabe S., Ahsan M.N., Funabara D.;
RT "Anchovy trypsinogen mRNA.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB041929; BAB40329.1; -.
DR HSSP; P00763; 1DPO.
DR MEROPS; S01.125; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.

```

```

DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM0020; TRYD_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease.
SQ SEQUENCE 240 AA; 26026 MW; 299B119BFF071464 CRC64;

Query Match 62.6%; Score 67; DB 13; Length 240;
Best Local Similarity 65.0%; Pred. No. 0.011; Mismatches 3; Gaps 0;
Matches 13; Conservative 3; Indels 4;

Qy 1 IVGGYTCGANTVPYQVSLNS 20
Db 20 IVGGYECQAHSQPHVSLNS 39

RESULT 12
Q9W7Q5 PRELIMINARY; PRT; 247 AA.
AC Q9W7Q5
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsinogen 3.
OS Paralicthys olivaceus (Flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidae; Paralicthidae; Paralicthys.
OX NCBI_TaxID=8255;
RN [1]
RP SEQUENCE FROM N.A.
RQ TISSUE=PANCREAS;
RA Suzuki T., Srivastava A.S., Kurokawa T.;
RT "Japanese flounder mRNA for trypsinogen 3.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB029752; BAA82364.2; -.
DR HSP; P00763; LDPO.
DR MEROPS; S01.124; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM0020; TRYD_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease.
FT NON_TER 1
SQ SEQUENCE 238 AA; 26071 MW; F2B8908085B8D062 CRC64;

Query Match 60.7%; Score 65; DB 13; Length 238;
Best Local Similarity 65.0%; Pred. No. 0.023; Mismatches 2; Indels 5; Gaps 0;
Matches 13; Conservative 2;

Qy 1 IVGGYTCGANTVPYQVSLNS 20
Db 18 IVGGYECTPVYQPHVSLNS 37

RESULT 14
Q98TG9 PRELIMINARY; PRT; 241 AA.
ID Q98TG9
AC Q98TG9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsinogen II.
GN ATRYII.
OS Engraulis japonicus (Japanese anchovy).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Clupeomorpha; Engraulidae;
OC Engraulis.
OX NCBI_TaxID=42892;
RN [1]
RP SEQUENCE FROM N.A.
RQ TISSUE=PYLORIC CAECA;
RA Watabe S., Ahsan M.N., Funabara D.;
RT "Anchovy trypsinogen mRNA.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB041930; BAB40330.1; -.
DR HSP; P00763; LDPO.
DR MEROPS; S01.258; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM0020; TRYD_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease.
SQ SEQUENCE 241 AA; 26282 MW; FE362D39CAEB2BF6 CRC64;

Query Match 60.7%; Score 65; DB 13; Length 241;
Matches 13; Conservative 2; Indels 5; Gaps 0;

```

```

DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM0020; TRYD_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease.
SQ SEQUENCE 240 AA; 26026 MW; 299B119BFF071464 CRC64;

Query Match 62.6%; Score 67; DB 13; Length 240;
Best Local Similarity 65.0%; Pred. No. 0.011; Mismatches 3; Gaps 0;
Matches 13; Conservative 3; Indels 4;

Qy 1 IVGGYTCGANTVPYQVSLNS 20
Db 20 IVGGYECQAHSQPHVSLNS 39

RESULT 12
Q9W7Q5 PRELIMINARY; PRT; 247 AA.
AC Q9W7Q5
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsinogen 3.
OS Paralicthys olivaceus (Flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidae; Paralicthidae; Paralicthys.
OX NCBI_TaxID=8255;
RN [1]
RP SEQUENCE FROM N.A.
RQ TISSUE=PANCREAS;
RA Suzuki T., Srivastava A.S., Kurokawa T.;
RT "Japanese flounder mRNA for trypsinogen 3.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB029752; BAA82364.2; -.
DR HSP; P00763; LDPO.
DR MEROPS; S01.124; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM0020; TRYD_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease.
SQ SEQUENCE 247 AA; 26948 MW; DC4B647179DD972 CRC64;

Query Match 61.7%; Score 66; DB 13; Length 247;
Best Local Similarity 63.2%; Pred. No. 0.017; Mismatches 3; Gaps 0;
Matches 12; Conservative 3; Indels 4;

Qy 1 IVGGYTCGANTVPYQVSLN 19
Db 22 IVGGYECQAHSQPHVSLN 40

RESULT 13
Q9W7Q6 PRELIMINARY; PRT; 238 AA.
AC Q9W7Q6
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsinogen 2 (Fragment).
OS Paralicthys olivaceus (Flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

Best Local Similarity 65.0%; Pred. No. 0.024;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20  
||||| : : : : :  
Db 20 IVGGYECQPYSQPHQVSLNS 39

## RESULT 15

Q9W7Q7 PRELIMINARY; PRT; 242 AA.  
AC Q9W7Q7;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Trypsinogen 1.  
OS Paraliichthys olivaceus (Flounder).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; pleuronectiformes;  
OC Pleuronectoidei; Paraliichthyidae; Paraliichthys.  
OX NCBI\_TaxID=8255;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PANCREAS;  
RA Suzuki T., Srivastava A.S., Kurokawa T.;  
RT "Japanese flounder mRNA for trypsinogen 1."  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
DR EMBL; AB029750; BAA82362.1; -.  
DR HSSP; P00763; LDPO.  
DR MEROPS; S01.125; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRYP\_SPG; 1.  
DR PROSITE; PS00134; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease  
SQ SEQUENCE 242 AA; 26548 MW; 6DA722C8BC194A2 CRC64;

Query Match 60.7%; Score 65; DB 13; Length 242;  
Best Local Similarity 65.0%; Pred. No. 0.024;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20  
||||| : : : : :  
Db 21 IVGGYECTPYSQPHQVSLNS 40

Search completed: February 12, 2003, 10:27:18  
Job time : 15.1493 secs

**THIS PAGE BLANK (usp)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:18:40 ; Search time 4.29851 Seconds  
(without alignments)  
136.898 Million cell updates/sec

Title: US-10-036-371-2

Perfect score: 107

Sequence: 1 IVGGYTCGANTVPYQVSLNS 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Issued first 45 summaries

Database :

Issued Patents AA: \*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	223	1	US-08-278-091-9
2	107	100.0	223	1	US-08-483-859-9
3	107	100.0	223	1	US-08-472-173-9
4	107	100.0	223	2	US-08-487-167-9
5	107	100.0	223	2	US-08-482-816-9
6	107	100.0	223	2	US-08-296-149-9
7	107	100.0	223	2	US-08-801-499-9
8	107	100.0	223	2	US-08-615-271-9
9	107	100.0	223	3	US-09-074-659-9
10	107	100.0	223	3	US-09-074-659-9
11	107	100.0	223	3	US-09-106-468-9
12	107	100.0	223	4	US-09-106-468A-9
13	107	100.0	223	4	US-09-106-467-9
14	107	100.0	223	3	US-09-120-582-2
15	96	89.7	247	2	US-08-956-267A-2
16	87	81.3	224	4	US-08-944-483-35
17	85	79.4	224	2	US-08-766-982-13
18	85	79.4	224	4	US-08-944-483-36
19	85	79.4	224	4	US-09-296-219-13
20	85	79.4	225	2	US-09-027-337-5
21	85	79.4	225	4	US-09-644-600-5
22	85	79.4	246	2	US-08-978-404B-44
23	84	78.5	281	4	US-08-944-483-34
24	68.5	64.0	281	1	US-08-467-155A-7
25	68.5	64.0	281	2	US-08-628-198-7
26	68.5	64.0	281	4	US-09-201-038-7
27	68.5	64.0	281	5	PCT-US96-07343-7

28	68	63.6	286	1	US-08-467-155A-9	Sequence 9, Appli
29	68	63.6	286	2	US-08-628-198-9	Sequence 9, Appli
30	68	63.6	286	4	US-09-201-038-9	Sequence 9, Appli
31	68	63.6	286	5	PCT-US96-07343-9	Sequence 9, Appli
32	67	62.6	25	2	US-08-385-540A-18	Sequence 18, Appl
33	67	62.6	25	2	US-08-600-273A-18	Sequence 18, Appl
34	67	62.6	25	3	US-08-486-820-18	Sequence 18, Appl
35	67	62.6	25	4	US-09-220-731-17	Sequence 17, Appl
36	64.5	60.3	299	1	US-08-467-155A-8	Sequence 8, Appli
37	64.5	60.3	299	2	US-08-628-198-8	Sequence 8, Appli
38	64.5	60.3	299	4	US-09-201-038-8	Sequence 8, Appli
39	64.5	60.3	299	5	PCT-US96-07343-8	Sequence 8, Appli
40	60	56.1	23	4	US-09-220-731-18	Sequence 18, Appl
41	60	56.1	25	2	US-08-385-540A-19	Sequence 19, Appl
42	60	56.1	25	2	US-08-385-540A-20	Sequence 20, Appl
43	60	56.1	25	2	US-08-600-273A-19	Sequence 19, Appl
44	60	56.1	25	2	US-08-600-273A-20	Sequence 20, Appl
45	60	56.1	25	3	US-08-486-820-19	Sequence 19, Appl

#### ALIGNMENTS

RESULT 1  
US-08-278-091-9  
; Sequence 9, Application US/08278091  
; Patent No. 5506139  
; GENERAL INFORMATION:  
; APPLICANT: LOOMORE, Sheena M  
; APPLICANT: YANG, Yan-Ping  
; APPLICANT: CHONG, Pele  
; APPLICANT: COMEN, Raymond P.  
; APPLICANT: KLEIN, Michael H.  
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with  
; TITLE OF INVENTION: Reduced Protease Activity  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: #1.0, Version #1.25  
; APPLICATION NUMBER: US/08/278,091  
; FILING DATE: 21-JUL-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1036-371  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 223 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-278-091-9

Query Match 100.0%; Score 107; DB 1; Length 223;  
Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20  
|||||

```
Db      1 IVGGYTCGANTVPYQVSLNS 20

RESULT 2
US-08-483-859-9
; Sequence 9, Application US/08483859
; Patent No. 5656436
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; REDUCED PROTEASE ACTIVITY
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,173
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-493 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-472-173-9

Query Match      100.0%; Score 107; DB 1; Length 223;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 IVGGYTCGANTVPYQVSLNS 20
      |||
Db      1 IVGGYTCGANTVPYQVSLNS 20
      |||

RESULT 4
US-08-487-167-9
; Sequence 9, Application US/08487167
; Patent No. 5869302
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; REDUCED PROTEASE ACTIVITY
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

Db      1 IVGGYTCGANTVPYQVSLNS 20
      |||
Qy      1 IVGGYTCGANTVPYQVSLNS 20
      |||

RESULT 3
US-08-472-173-9
; Sequence 9, Application US/08472173
; Patent No. 566353
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,167
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-508 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-487-167-9

Query Match 100.0%; Score 107; DB 2; Length 223;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20
Db 1 IVGGYTCGANTVPYQVSLNS 20

RESULT 5
US-08-482-816-9
; Sequence 9, Application US/08482816
; Patent No. 5935573
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Fele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced Protease Act
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-494 MIS:vg
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-482-816-9

Query Match 100.0%; Score 107; DB 2; Length 223;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20
Db 1 IVGGYTCGANTVPYQVSLNS 20

RESULT 6
US-08-296-149-9
; Sequence 9, Application US/08296149
; Patent No. 5939297
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Fele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-390
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-296-149-9

Query Match 100.0%; Score 107; DB 2; Length 223;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20
Db 1 IVGGYTCGANTVPYQVSLNS 20

RESULT 7

```

```
US-08-801-499-9
; Sequence 9, Application US/08801499
; Patent No. 5962430
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced Protease Act
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,499
; FILING DATE: 26-AUG-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/482,816
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-671 MIS:jb
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-801-499-9

Query Match 100.0%; Score 107; DB 2; Length 223;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IVGGYTCGANTVPYQVSLNS 20
Db 1 IVGGYTCGANTVPYQVSLNS 20
|||||
|||||

RESULT 8
US-08-615-271-9
; Sequence 9, Application US/08615271
; Patent No. 5981503
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,660
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,167
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/615,271
; FILING DATE: 20-JUN-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-580
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-615-271-9

Query Match 100.0%; Score 107; DB 2; Length 223;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IVGGYTCGANTVPYQVSLNS 20
Db 1 IVGGYTCGANTVPYQVSLNS 20
|||||
|||||

RESULT 9
US-09-074-660-9
; Sequence 9, Application US/09074660
; Patent No. 6020183
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,660
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,167
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
```



; APPLICATION NUMBER: US 08/296,149  
; FILING DATE: 26-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/278,091  
; FILING DATE: 21-JUL-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-731 MIS.jb  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 223 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-074-660-9

Query Match 100.0%; Score 107; DB 3; Length 223;  
Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20  
|||||  
DB 1 IVGGYTCGANTVPYQVSLNS 20

## RESULT 10

US-09-074-659-9  
; Sequence 9, Application US/09074659  
; Patent No. 6025342

## GENERAL INFORMATION:

; APPLICANT: LOOSMORE, Sheena M.  
; APPLICANT: YANG, Yan-Ping  
; APPLICANT: CHONG, Pele  
; APPLICANT: COMEN, Raymond P.  
; APPLICANT: KLEIN, Michel H.  
; TITLE OF INVENTION: Analog of Haemophilus Hin47 protein with  
; TITLE OF INVENTION: Reduced Protease Activity  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/074,659  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/487,167  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/296,149  
; FILING DATE: 26-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/278,091  
; FILING DATE: 21-JUL-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-730 MIS.jb  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 223 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-074-659-9

Query Match 100.0%; Score 107; DB 3; Length 223;  
Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20  
|||||  
DB 1 IVGGYTCGANTVPYQVSLNS 20

## RESULT 11

US-09-106-468-9  
; Sequence 9, Application US/09106468  
; Patent No. 6114125

## GENERAL INFORMATION:

; APPLICANT: LOOSMORE, Sheena M.  
; APPLICANT: YANG, Yan-Ping  
; APPLICANT: CHONG, Pele  
; APPLICANT: COMEN, Raymond P.  
; APPLICANT: KLEIN, Michel H.  
; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED  
; TITLE OF INVENTION: PROTEASE ACTIVITY  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/106,468  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/615,271  
; FILING DATE: 20-JUN-1996  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-825  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 223 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-106-468-9

Query Match 100.0%; Score 107; DB 3; Length 223;  
Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20  
|||||  
DB 1 IVGGYTCGANTVPYQVSLNS 20

```
RESULT 12
US-09-106-466A-9
; Sequence 9, Application US/09106466A
; Patent No. 6147057
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,466A
; FILING DATE:
; CLASSIFICATION: 514514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/615,271
; FILING DATE: 20-JUN-1996
; CLASSIFICATION: 514514
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-826
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-106-466A-9
;
; Query Match 100.0%; Score 107; DB 4; Length 223;
; Best Local Similarity 100.0%; Pred. No. 1.5e-08;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IVGGYTCGANTVPYQVSLNS 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 IVGGYTCGANTVPYQVSLNS 20
| | | | | | | | | | | | | | | | | | | | | |

RESULT 13
US-09-106-467-9
; Sequence 9, Application US/09106467
; Patent No. 6153580
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,467
; FILING DATE:
; CLASSIFICATION: 514514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/615,271
; FILING DATE: 20-JUN-1996
; CLASSIFICATION: 514514
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-826
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-106-466A-9
;
; Query Match 100.0%; Score 107; DB 4; Length 223;
; Best Local Similarity 100.0%; Pred. No. 1.5e-08;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IVGGYTCGANTVPYQVSLNS 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 IVGGYTCGANTVPYQVSLNS 20
| | | | | | | | | | | | | | | | | | | | | |

RESULT 14
US-09-120-582-2
; Sequence 2, Application US/09120582
; Patent No. 6087558
; GENERAL INFORMATION:
; APPLICANT: Howard, John A.
; APPLICANT: Hood, Elizabeth
; TITLE OF INVENTION: COMMERCIAL PRODUCTION OF PROTEASES IN PLANTS
; FILE REFERENCE: 10014
; CURRENT APPLICATION NUMBER: US/09/120,582
; CURRENT FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Bos taurus
;
US-09-120-582-2
;
; Query Match 100.0%; Score 107; DB 3; Length 229;
; Best Local Similarity 100.0%; Pred. No. 1.5e-08;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IVGGYTCGANTVPYQVSLNS 20
| | | | | | | | | | | | | | | | | | | | | |
Db 7 IVGGYTCGANTVPYQVSLNS 26
| | | | | | | | | | | | | | | | | | | | | |

RESULT 15
US-08-956-267A-2
```

```
; Sequence 2, Application US/08956267A
; Patent No. 5945328
; GENERAL INFORMATION:
; APPLICANT: WOLDIKE, Helle Fabricius
; APPLICANT: KJELDSSEN, Thomas Borglum
; TITLE OF INVENTION: A Process For Producing Trypsin
; TITLE OF INVENTION: (Trypsinogen)
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5945328o No. 5945328disk of No. 5945328th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,267A
; FILING DATE: 22-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 4500.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 247 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-956-267A-2
```

```
Query Match      89.7%; Score 96; DB 2; Length 247;
Best Local Similarity 85.0%; Pred. No. 6.9e-07;
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
Oy 1 IVGGYTCGANTVPYQVSLNS 20
Db 25 IVGGYTCGANTVPYQVSLNS 44
```

```
Search completed: February 12, 2003, 10:30:06
Job time : 4.29851 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:22:36 ; Search time 3.04478 Seconds  
(without alignments)  
167.821 Million cell updates/sec

Title: US-10-036-371-2

Perfect score:

Sequence: 1 IVGGYTCGANTVPYQVSLNS 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length:	200000000
Maximum DB seq length:	200000000

Post-processing: Minimum Match 0%

Loss processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA: \*

- 1: /cgn2\_6/ptodata1/1/pubpaa/us08\_NEW\_PUB.pcp:
- 2: /cgn2\_6/ptodata1/1/pubpaa/PCT\_NEW\_PUB.pcp:
- 3: /cgn2\_6/ptodata1/1/pubpaa/US10\_NEW\_PUB.pcp:
- 4: /cgn2\_6/ptodata1/1/pubpaa/US06\_PUBCOMB.pcp:
- 5: /cgn2\_6/ptodata1/1/pubpaa/US07\_NEW\_PUB.pcp:
- 6: /cgn2\_6/ptodata1/1/pubpaa/US07\_PUBCOMB.pcp:
- 7: /cgn2\_6/ptodata1/1/pubpaa/PCTUS\_PUBCOMB.pcp:
- 8: /cgn2\_6/ptodata1/1/pubpaa/us08\_PUBCOMB.pcp:
- 9: /cgn2\_6/ptodata1/1/pubpaa/us09\_NEW\_PUB.pcp:
- 10: /cgn2\_6/ptodata1/1/pubpaa/us09\_PUBCOMB.pcp:
- 11: /cgn2\_6/ptodata1/1/pubpaa/US10\_NEW\_PUB.pcp:
- 12: /cgn2\_6/ptodata1/1/pubpaa/US10\_PUBCOMB.pcp:
- 13: /cgn2\_6/ptodata1/1/pubpaa/US60\_NEW\_PUB.pcp:
- 14: /cgn2\_6/ptodata1/1/pubpaa/US60\_PUBCOMB.pcp:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES									
Result No.	Score	Query #		Length	DB	ID	Description		
		Match							
1	107	100.0	20	12	US-10-036	171-2	Sequence 2, Appli		
2	98	91.6	223	10	US-09-010	771-14	Sequence 14, Appl		
3	84	78.5	247	10	US-03-923	773-154	Sequence 154, App		
4	68.5	64.0	281	12	US-10-021	368-7	Sequence 7, Appli		
5	68	63.6	286	12	US-10-021	368-9	Sequence 9, Appli		
6	67	62.6	25	9	US-09-938	369-14	Sequence 14, Appl		
7	65	60.7	37	9	US-09-938	369-17	Sequence 17, Appl		
8	64.5	60.3	299	12	US-10-021	368-8	Sequence 8, Appli		
9	63	58.9	20	12	US-10-036	371-1	Sequence 1, Appli		
10	61	57.0	225	12	US-10-036	371-8	Sequence 8, Appli		
11	60	56.1	25	9	US-09-938	369-15	Sequence 15, Appl		
12	60	56.1	25	9	US-09-938	369-16	Sequence 16, Appl		
13	58	54.2	271	12	US-10-021	368-10	Sequence 10, Appl		
14	50	46.7	1128	10	US-09-888	615-97	Sequence 97, Appl		
15	49	45.8	235	10	US-03-925	300-1093	Sequence 1093, Ap		
16	48	44.9	263	10	US-03-764	762-5	Sequence 5, Appli		
17	48	44.9	263	10	US-09-988	975A-8	Sequence 8, Appli		
18	44	44.9	311	10	US-09-900	754-2	Sequence 2, Appli		
19	46	43.0	261	9	US-09-895	793-946	Sequence 946, App		

Sequence 62, Appl  
Sequence 10, Appl  
Sequence 946, App  
Sequence 6, Appl  
Sequence 1, Appl  
Sequence 4, Appl  
Sequence 7, Appl  
Sequence 745, App  
Sequence 11, Appl  
Sequence 617, App  
Sequence 617, App  
Sequence 617, App  
Sequence 617, App  
Sequence 617, App  
Sequence 617, App  
Sequence 978, App  
Sequence 978, App  
Sequence 982, App  
Sequence 982, App  
Sequence 947, App  
Sequence 947, App  
Sequence 91, Appl  
Sequence 11, Appl  
Sequence 10, Appl  
Sequence 13875, A  
Sequence 19, Appl

## ALIGNMENTS

```

RESULT 1
US-10-036-371-2
; Sequence 2, Application US/10036371
; Patent No. US20020141987A1
; GENERAL INFORMATION:
; APPLICANT: BVARNARSON, JON B.
; TITLE OF INVENTION: FISH SERINE PROTEINASES AND THEIR PHARMACEUTICAL AND
; FILE REFERENCE: 81691/284960
; CURRENT APPLICATION NUMBER: US/10/036,371
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: 09/411,688
; PRIOR FILING DATE: 1999-10-12
; PRIOR APPLICATION NUMBER: 5086/99
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Bovine sp.
US-10-036-371-2

```

```
; TITLE OF INVENTION: Structures of Molecules
; FILE REFERENCE: 522.1921D2
; CURRENT APPLICATION NUMBER: US/09/910,071
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: US 08/014,867
; PRIOR FILING DATE: 1993-02-08
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 223
; TYPE: PRT
; ORGANISM: unknown
; FEATURE:
; OTHER INFORMATION: amino acid sequence of trypsin
US-09-910-071-14

Query Match          91.6%; Score 98; DB 10; Length 223;
Best Local Similarity 95.0%; Pred. No. 1.6e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IVGGYTCGANTVPYQVSLNS 20
Db 1 IVGGYTCANTVPYQVSLNS 20

RESULT 3
US-09-923-779-154
; Sequence 154, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 154
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-923-779-154

Query Match          78.5%; Score 84; DB 10; Length 247;
Best Local Similarity 80.0%; Pred. No. 2.2e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 IVGGYTCGANTVPYQVSLNS 20
Db 24 IVGGYICEENSVPYQVSLNS 43

RESULT 4
US-10-021-368-7
; Sequence 7, Application US/10021368
; Patent No. US20020106367A1
; GENERAL INFORMATION:
; APPLICANT: Band, Vimla
; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED
; MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; MEDIUM TYPE: Floppy disk
; COMPUTER READABLE FORM:
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/10/021,368
; FILING DATE: 12-Dec-2001
; CLASSIFICATION: <Unknown>
; APPLICATION DATA:
; APPLICATION NUMBER: 09/201,038
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00398/100002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/10/021,368
; FILING DATE: 12-Dec-2001
; CLASSIFICATION: <Unknown>
; APPLICATION DATA:
; APPLICATION NUMBER: 09/201,038
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00398/100002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; MEDIUM TYPE: Floppy disk
; COMPUTER READABLE FORM:
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/10/021,368
; FILING DATE: 12-Dec-2001
; CLASSIFICATION: <Unknown>
; APPLICATION DATA:
; APPLICATION NUMBER: 09/201,038
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00398/100002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; MEDIUM TYPE: Floppy disk

Query Match          64.0%; Score 68.5; DB 12; Length 281;
Best Local Similarity 55.6%; Pred. No. 0.0049;
Matches 15; Conservative 3; Mismatches 2; Indels 7; Gaps 1;

Qy 1 IVGGY-----TCGANTVPYQVSLNS 20
Db 24 IVGGYMMTRYARTCRSSVPYQVSLNA 50

RESULT 5
US-10-021-368-9
; Sequence 9, Application US/10021368
; Patent No. US20020106367A1
; GENERAL INFORMATION:
; APPLICANT: Band, Vimla
; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED
; MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; MEDIUM TYPE: Floppy disk
; COMPUTER READABLE FORM:
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/10/021,368
; FILING DATE: 12-Dec-2001
; CLASSIFICATION: <Unknown>
; APPLICATION DATA:
; APPLICATION NUMBER: 09/201,038
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00398/100002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; MEDIUM TYPE: Floppy disk
```

```
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 286 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-10-021-368-9

Query Match      63.6%; Score 68; DB 12; Length 286;
Best Local Similarity 53.6%; Pred. No. 0.0059;
Matches 15; Conservative 3; Mismatches 2; Indels 8; Gaps 1;

QY 1 IVGGY-----TCGANTVPYQVSLNS 20
Db 25 IVGGYRNYRYPVBTQEHSPYQVSLNA 52

RESULT 6
US-09-938-269-14
; Sequence 14, Application US/09938269
; Publication No. US20030007951A1
; GENERAL INFORMATION:
; APPLICANT: Franklin, Richard L.
; APPLICANT: Cowling, Didier S.P.
; APPLICANT: Hubbel, Jeffrey A.
; APPLICANT: van de Wetering, Petra
; TITLE OF INVENTION: Treatment of Trauma
; FILE REFERENCE: 314572-105
; CURRENT APPLICATION NUMBER: US/09/938,269
; CURRENT FILING DATE: 2001-08-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Salmon enzyme
US-09-938-269-14

Query Match      62.6%; Score 67; DB 9; Length 25;
Best Local Similarity 70.0%; Pred. No. 0.0067;
Matches 14; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVGGYTCGANTVPYQVSLNS 20
Db 1 IVGGYCKAYSQAYQVSLNS 20

RESULT 7
US-09-938-269-17
; Sequence 17, Application US/09938269
; Publication No. US20030007951A1
; GENERAL INFORMATION:
; APPLICANT: Franklin, Richard L.
; APPLICANT: Cowling, Didier S.P.
; APPLICANT: Hubbel, Jeffrey A.
; APPLICANT: van de Wetering, Petra
; TITLE OF INVENTION: Treatment of Trauma
; FILE REFERENCE: 314572-105
; CURRENT APPLICATION NUMBER: US/09/938,269
; CURRENT FILING DATE: 2001-08-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Atlantic cod
US-09-938-269-17

Query Match      60.7%; Score 65; DB 9; Length 37;
Best Local Similarity 65.0%; Pred. No. 0.002;
Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 286 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-10-021-368-9

Query Match      60.3%; Score 64.5; DB 12; Length 299;
Best Local Similarity 57.1%; Pred. No. 0.02;
Matches 16; Conservative 2; Mismatches 1; Indels 9; Gaps 2;

QY 1 IVGGY-----TCGANTVPYQVSLNS 20
Db 37 IVGGYHSTRYIVBTC-ENSLPYQVSLNS 63

RESULT 9
US-10-036-371-1
; Sequence 1, Application US/10036371
; Patent No. US20020141987A1
; GENERAL INFORMATION:
; APPLICANT: BJARNARSON, JON B.
; TITLE OF INVENTION: FISH SERINE PROTEINASES AND THEIR PHARMACEUTICAL AND
; FILE REFERENCE: 81691/284960
; TITLE OF INVENTION: COSMETIC USE
; CURRENT APPLICATION NUMBER: US/10/036,371
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: 09/411,688
```

```

; PRIOR FILING DATE: 1999-10-12
; PRIOR APPLICATION NUMBER: 5086/99
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Gadus sp.
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (6)
; OTHER INFORMATION: Q or E
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (8)
; OTHER INFORMATION: E or T
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (9)
; OTHER INFORMATION: K or R
; US-10-036-371-1

```

```

Query Match      58.9%; Score 63; DB 12; Length 20;
Best Local Similarity 60.0%; Pred. No. 0.0021;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy 1 IVGGYTCGANTVPYQVSLNS 20
||||| | : : |||||
Db 1 IVGGYCXHXSHQAHOVSLNS 20

```

```

RESULT 10
US-10-036-371-8
; Sequence 8, Application US/10036371
; Patent No. US20020141987A1
; GENERAL INFORMATION:
; APPLICANT: BJARNARSON, JON B.
; TITLE OF INVENTION: FISH SERINE PROTEINASES AND THEIR PHARMACEUTICAL AND
; FILE REFERENCE: COSMETIC USE
; FILE REFERENCE: 81691/284960
; CURRENT APPLICATION NUMBER: US/10/036,371
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: 09/411,688
; PRIOR FILING DATE: 1999-10-12
; PRIOR APPLICATION NUMBER: 5086/99
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Gadus sp.
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (9)
; OTHER INFORMATION: K or R
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (24)
; OTHER INFORMATION: Y or F
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (32)
; OTHER INFORMATION: K or E
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (33)
; OTHER INFORMATION: D or Q
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (64)
; OTHER INFORMATION: Y or F

```

```

; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (71)
; OTHER INFORMATION: I or unknown
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (82)
; OTHER INFORMATION: N or D
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (92)
; OTHER INFORMATION: K or E
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (129)
; OTHER INFORMATION: A or D
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (136)
; OTHER INFORMATION: V or C
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (138)
; OTHER INFORMATION: N or S
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (204)
; OTHER INFORMATION: H or N
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (212)
; OTHER INFORMATION: M, V or C
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (224)
; OTHER INFORMATION: N or S
; US-10-036-371-8

```

```

Query Match      57.0%; Score 61; DB 12; Length 225;
Best Local Similarity 60.0%; Pred. No. 0.05;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy 1 IVGGYTCGANTVPYQVSLNS 20
||||| | : : |||||
Db 1 IVGGYECTXHSQAHOVSLNS 20

```

```

RESULT 11
US-09-938-269-15
; Sequence 15, Application US/09938269
; Publication No. US20030007951A1
; GENERAL INFORMATION:
; APPLICANT: Franklin, Richard L.
; APPLICANT: Cowling, Didier S.P.
; APPLICANT: Hubbel, Jeffrey A.
; APPLICANT: van de Wetering, Petra
; TITLE OF INVENTION: Treatment of Trauma
; FILE REFERENCE: 314572-105
; CURRENT APPLICATION NUMBER: US/09/938,269
; CURRENT FILING DATE: 2001-08-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Atlantic cod
; US-09-938-269-15

```

```

Query Match      56.1%; Score 60; DB 9; Length 25;
Best Local Similarity 60.0%; Pred. No. 0.0073;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy 1 IVGGYTCGANTVPYQVSLNS 20

```



```

RESULT 15
; US-09-925-300-1093
; Sequence 1093, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05968
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1093
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-925-300-1093

```

Query Match 45.8%; Score 49; DB 10; Length 235;  
Best Local Similarity 45.0%; Pred. No. 3.1;  
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 IVGGYTCGANTVPYQVSLNS 20  
Db 87 IVGWECEKHSQPMQVAVIS 106

Search completed: February 12, 2003, 10:31:04  
Job time : 4.04478 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:03:55 ; Search time 8.73134 Seconds  
(without alignments)  
198.395 Million cell updates/sec

Title: US-10-036-371-3  
Perfect score: 59  
Sequence: 1 CGXPAIXPXXTG

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

```
Minimum DB seq length: 0
Maximum DB seq length: 20
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```

Database :
A. Geneseq_101002.*
1: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1980.DAT.*
2: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1981.DAT.*
3: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1982.DAT.*
4: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1983.DAT.*
5: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1984.DAT.*
6: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1985.DAT.*
7: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1986.DAT.*
8: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1987.DAT.*
9: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1988.DAT.*
10: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1989.DAT.*
11: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1990.DAT.*
12: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1991.DAT.*
13: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1992.DAT.*
14: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1993.DAT.*
15: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1994.DAT.*
16: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1995.DAT.*
17: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1996.DAT.*
18: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1997.DAT.*
19: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1998.DAT.*
20: /SID52/cgcdata/geneseq/geneseqp_emb1/AA1999.DAT.*
21: /SID52/cgcdata/geneseq/geneseqp_emb1/AA2000.DAT.*
22: /SID52/cgcdata/geneseq/geneseqp_emb1/AA2001.DAT.*
23: /SID52/cgcdata/geneseq/geneseqp_emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	51	86.4	13	22	AAB31575	Amino terminal seq
2	40	67.8	13	20	AAV50209	Neutrophil-activat
3	40	67.8	13	22	AAB31577	Amino terminal seq
4	40	67.8	15	20	AAV50207	Neutrophil-activat
5	40	67.8	146	21	AAB54191	Human pancreatic c
6	40	67.8	245	21	AAV99596	Bovine chymotrypsi
7	40	67.8	263	23	AAU82738	Amino acid sequenc
8	39	66.1	15	20	AAV50208	Neutrophil-activat
9	39	66.1	15	20	AAV50210	Neutrophil-activat
10	39	66.1	811	12	AAR12944	Plasminogen mutain

11	39	66.1	813	12	AA812936
12	39	66.1	813	12	AA812937
13	38	64.4	626	22	AB861150
14	36	61.0	51	22	AAU54358
15	36	61.0	106	23	ABF09202
16	36	61.0	264	21	AA811711
17	36	61.0	264	21	AAV50212
18	35	59.3	13	20	AAV50212
19	35	59.3	44	21	AAV64742
20	35	59.3	44	23	ABF03011
21	35	59.3	88	22	AAU61671
22	35	59.3	192	21	AA854077
23	35	59.3	264	21	AA811710
24	35	59.3	801	22	AA803738
25	35	59.3	887	20	AAV01711
26	35	59.3	885	23	AAU81956
27	35	59.3	936	22	AAE03735
28	35	59.3	948	22	AAE03736
29	35	59.3	948	23	AAU75168
30	35	59.3	954	19	AAW48978
31	35	59.3	954	20	AAV17821
32	35	59.3	954	21	AA801312
33	35	59.3	954	23	AAU86132
34	35	59.3	955	22	AAE12889
35	35	59.3	962	22	AAE03734
36	34	57.6	110	21	AA840871
37	34	57.6	110	23	ABP32707
38	34	57.6	112	21	AAE03081
39	34	57.6	117	21	AAE60978
40	34	57.6	117	21	AAE60980
41	34	57.6	120	22	AA896645
42	34	57.6	121	21	AAAG04376
43	34	57.6	121	21	AAE60977
44	34	57.6	121	21	AAE60979
45	34	57.6	183	21	AAAG04375
46	34	57.6	183	21	AAE60976

## ALIGNMENTS

RESULT 1	
AAB31575	
ID	AAB31575 standard; peptide; 13 AA.
XX	
XX	AAB31575;
XX	AC
DT	20-APR-2001 (first entry)
XX	
DE	Amino terminal sequence of cod try





XX OS Homo sapiens.  
 XX PN WO200055320-A1.  
 XX PD 21-SEP-2000.  
 XX PF 08-MAR-2000; 2000WO-US05989.  
 XX PR 12-MAR-1999; 99US-0124270.  
 XX PA (HUMA-) HUMAN GENOME SCI INC.  
 XX PI Rosen CA, Ruben SM;  
 DR WPI; 2000-579444/54.  
 DR N-PSDB; AAC98956.  
 XX PT New nucleic acid that is a pancreatic cancer antigen for preventing,  
 PT treating, or ameliorating a medical condition, particular pancreatic  
 PT cancer, or for use in assays for diagnosing a pathological condition -  
 XX  
 PS Claim 11; Page 1081; 1379pp; English.  
 XX  
 CC AAC98773 to AAC99231 encode the human pancreatic cancer associated  
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to  
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,  
 CC neuroprotective, cardioprotective, immunomodulatory, relaxant, contraceptive,  
 CC gynaecological, nontoxic and antiinflammatory activities, and can be used  
 CC in gene therapy. The polynucleotide and proteins can be used for  
 CC preventing, treating, or ameliorating a medical condition or in assays  
 CC for diagnosing a pathological condition or a susceptibility to one in a  
 CC subject. Binding partners to the proteins and the activity of the  
 CC proteins can be identified. The pancreatic cancer antigens can be used to  
 CC detect, treat or prevent pancreatic disorders, especially cancer.  
 CC Agonists and antagonists to the antigens can be screened for. The  
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
 CC acid hybridization probes that can be used in chromosome mapping, linkage  
 CC analysis, tissue identification and/or typing and a variety of forensic  
 CC and diagnostic methods. The proteins can be used to generate antibodies  
 CC which are used to purify, detect and target the polypeptides, including  
 CC both in vivo and in vitro diagnostic and therapeutic methods. The  
 CC proteins can be used to treat or prevent neural, immune system, muscular,  
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
 CC sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 146 AA;  
 Query Match 67.8%; Score 40; DB 21; Length 146;  
 Best Local Similarity 58.3%; Pred. No. 11;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 Qy 1 CGXPAIXPXXTG 12  
 |||||  
 Db 23 CGVPAIHPVLSG 34  
 RESULT 6  
 AAY99596  
 ID AAY99596 standard; protein; 245 AA.  
 XX  
 AC AAY99596;  
 XX  
 DT 13-SEP-2000 (first entry)  
 XX  
 DE Bovine chymotrypsinogen A.  
 XX  
 KW Bovine; plasminogen activator; cardiant; thrombolytic;  
 KW heart attack; stroke; blood clotting disorder.  
 XX  
 OS Bos taurus.  
 XX

PN WO200032759-A1.  
 XX  
 PD 08-JUN-2000.  
 XX  
 PF 06-MAY-1999; 99WO-US09991.  
 XX  
 PR 02-DEC-1998; 98US-0110588.  
 XX  
 PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 XX  
 PI Lin X, Zhang XC, Tang JUN;  
 XX  
 DR WPI; 2000-422975/36.  
 XX  
 PT Polypeptide with plasminogen activator activity useful as thrombolytic  
 PT agent for treating blood clot disorders e.g. heart attack, comprises 10  
 PT amino acid peptide fragment for recognition or activation of  
 PT plasminogen -  
 XX  
 PS Disclosure; Page 35-36; 41pp; English.  
 XX  
 CC The present sequence is bovine chymotrypsinogen. It was included in  
 CC a review of sequence homologies of several plasminogen activators.  
 CC Plasminogen is the principal serine protease zymogen in the  
 CC extracellular fluids of vertebrates. Its active form, plasmin, is  
 CC implicated in pericellular proteolysis associated with a wide range of  
 CC physiological and pathological processes. Plasminogen expression is  
 CC regulated by plasminogen activators which hydrolyse a peptide bond in  
 CC plasminogen to convert it to plasmin or form tight binding complexes  
 CC with plasminogen to spontaneously convert it to plasmin. The sequence  
 CC homology analysis has identified a six amino acid peptide involved in  
 CC plasminogen activation. This peptide is particularly useful when inserted  
 CC between amino acid residues 644 and 645 of full length human  
 CC plasminogen. Novel plasminogen activators have been made based upon the  
 CC plasminogen activation/recognition site of plasminogen binding proteins.  
 CC The polypeptides are useful in preparing thrombolytic agents for treating  
 CC blood clotting disorders such as heart attack.  
 XX  
 SQ Sequence 245 AA;  
 Query Match 67.8%; Score 40; DB 21; Length 245;  
 Best Local Similarity 58.3%; Pred. No. 19;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 Qy 1 CGXPAIXPXXTG 12  
 |||||  
 Db 1 CGVPAIQPVLSG 12  
 RESULT 7  
 AAU82738  
 ID AAU82738 standard; Protein; 263 AA.  
 XX  
 AC AAU82738;  
 XX  
 DT 23-APR-2002 (first entry)  
 XX  
 DE Amino acid sequence of novel human protease #37.  
 XX  
 KW Human; protease; cancer; immune-related disorder; cardiovascular disease;  
 KW neuronal-associated disease; metabolic disorder; inflammatory disorder;  
 KW nervous system disorder; sexual dysfunction; pain; mood disorder;  
 KW hypertension; psychotic disorder; neurological disorder; dyskinesia;  
 KW viral infection; human immunodeficiency virus; HIV; non-viral infection;  
 KW ocular disease; cytostatic; enzyme.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200200860-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 26-JUN-2001; 2001WO-US20171.



XX 11-MAR-1999; 99WO-US05247.  
 XX 11-MAR-1998; 98US-0038894.  
 XX (CELL-) CELL ACTIVATION INC.  
 PA (REGC ) UNIV CALIFORNIA.  
 PA (SCRI ) SCRIPPS RES INST.  
 XX Stoughton RB, Schmid-Schonbein GW, Hugli TE, Kistler E;  
 XX WPI; 1999-580234/49.  
 XX Use of cell activating compositions in developing products for  
 PT diagnosis and treatment of e.g. cardiovascular, inflammatory,  
 PT autoimmune or Alzheimer's disease, trauma, arthritis, organ rejection,  
 PT diabetes, stroke or ischemia -  
 XX Example 9; Page 182; 184pp; English.  
 XX This invention describes a novel method for the use and preparation of  
 CC cell activating compositions which involves preparing a cell activating  
 CC composition comprising (a) homogenizing pancreatic tissue in buffer at  
 CC about neutral or higher pH to produce a homogenate; (b) removing  
 CC particulates from the homogenate; (c) optionally incubating the  
 CC resulting homogenate, with particulates removed, with a protease; and  
 CC (d) fractionating the homogenate and selecting fractions that exhibit  
 CC cell activation activity. The methods can be used for improving  
 CC treatment outcome or reducing risk of treatment of e.g. cardiovascular  
 CC disease, inflammatory disease, trauma, autoimmune diseases, arthritis,  
 CC organ rejection, diabetes and diabetic complications, stroke, ischemia,  
 CC Alzheimer's disease, myocardial infarction, haemorrhagic shock, diabetic  
 CC retinopathy, diabetes, venous insufficiency, unstable angina or trauma.  
 CC They can be used in the veterinary treatment of a non-human subject.  
 CC Protease inhibitors can be used to lower cell activation resulting from  
 CC these diseases and deficiencies. The detection of an elevated level of  
 CC hydrogen peroxide can be used to detect an inflammatory condition. An  
 CC elevated level of hydrogen peroxide in plasma or whole blood and in the  
 CC presence of superoxide dismutase (SOD) indicates leukocyte up  
 CC regulation, e.g. indicative of the onset of an acute cardiovascular  
 CC disorders, such as disease onset or ischemic complications. An elevated  
 CC level of hydrogen peroxide in plasma or whole blood and a low level in  
 CC the presence of SOD is indicative of a chronic or immune compromised  
 CC condition e.g. hypertension or sepsis. AAY50201-Y50334 represent peptides  
 CC used in the method of the invention.  
 XX SQ Sequence 15 AA;

Query Match 66.1%; Score 39; DB 20; Length 15;  
 Best Local Similarity 58.3%; Pred. No. 1.8;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CGXPAIXPXXTG 12  
 |||||  
 Db 1 CGVPAIPVLSG 12

RESULT 10  
 AAR12944  
 ID AAR12944 standard; Protein; 811 AA.  
 XX AAR12944;  
 AC AAR12944;  
 XX 23-SEP-1991 (first entry)  
 XX Plasminogen mutein T14 with thrombin cleavage site.  
 DE protease; fibrinolysis; blood clotting; thrombosis.  
 XX Key Location/Qualifiers  
 XX Cleavage-site 579..583  
 FT /note= "recognised by thrombin"  
 PT

PN WO9109118-A.  
 XX 27-JUN-1991.  
 XX 07-DEC-1990; 90WO-GB01912.  
 XX 07-DEC-1989; 89GB-0027722.  
 PR 07-DEC-1990; 90WO-GB01911.  
 XX (BRBI-) BRIT BIO-TECHN LTD.  
 XX Dawson KM, Edwards RM, Forman JM;  
 XX WPI; 1991-208145/28.  
 DR N-PSDB; AAQ12553.  
 XX Activatable fibrinolytic and antithrombotic proteins - activated by  
 PT e.g. factor Xa, thrombin or activated protein C  
 XX Claim 9; Fig 2 and Fig 5; 73pp; English.  
 XX This protein is a plasminogen mutant cleavable by thrombin.  
 CC Activation is localised to the thrombus because cleavage to plasmin  
 CC is by an enzyme of the blood clotting pathway. Compositions  
 CC comprising the mutant plasminogen are used for treatment or  
 CC prevention of thrombosis, etc.  
 CC See AAQ12542-Q12558.  
 XX SQ Sequence 811 AA;  
 Query Match 66.1%; Score 39; DB 12; Length 811;  
 Best Local Similarity 46.2%; Pred. No. 92;  
 Matches 6; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 Qy 1 CGXPAIXPXXTG 13  
 |||||  
 Db 567 CGRPQVSPKCGY 579  
 RESULT 11  
 AAR12936  
 ID AAR12936 standard; Protein; 813 AA.  
 XX AAR12936;  
 AC AAR12936;  
 XX 23-SEP-1991 (first entry)  
 XX Plasminogen mutein X5 with factor Xa cleavage site.  
 DE protease; fibrinolysis; blood clotting; thrombosis.  
 XX Key Location/Qualifiers  
 XX Cleavage-site 578..583  
 FT /label= Factor Xa cleavage site  
 XX WO9109118-A.  
 XX 27-JUN-1991.  
 XX 07-DEC-1990; 90WO-GB01912.  
 XX 07-DEC-1989; 89GB-0027722.  
 PR 07-DEC-1990; 90WO-GB01911.  
 XX (BRBI-) BRIT BIO-TECHN LTD.  
 XX Dawson KM, Edwards RM, Forman JM;  
 XX WPI; 1991-208145/28.  
 DR N-PSDB; AAQ12545.  
 XX Activatable fibrinolytic and antithrombotic proteins - activated by  
 PT e.g. factor Xa, thrombin or activated protein C



```
XX PS Claim 7; Fig 2 and Fig 4; 73pp; English.
XX CC This protein is a plasminogen mutant cleavable by Factor Xa.
XX CC Activation is localised to the thrombus because cleavage to plasmin
XX CC is by an enzyme of the blood clotting pathway. Compositions
XX CC comprising the mutant plasminogen are used for treatment or
XX CC prevention of thrombosis, etc.
XX CC See AAQ12542-Q12558.
XX SQ Sequence 813 AA;
XX
XX Query Match 66.1%; Score 39; DB 12; Length 813;
XX Best Local Similarity 46.2%; Pred. No. 92;
XX Matches 6; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
XX
QY 1 CGXPAIXPXXTGY 13
XX ||| : |||
Db 567 CGKPQVEPKKGY 579

RESULT 12
AAR12937
ID AAR12937 standard; Protein; 813 AA.
XX
XX AC AAR12937;
XX
XX DT 23-SEP-1991 (first entry)
XX
XX DE Plasminogen mutcin X6 with factor Xa cleavage site.
XX
XX KW protease; fibrinolysis; blood clotting; thrombosis.
XX
XX FH Key Location/Qualifiers
XX FT Cleavage-site 578..583
XX FT /label= Factor Xa cleavage site
XX
XX PN W09109118-A.
XX
XX PD 27-JUN-1991.
XX
XX PF 07-DEC-1990; 90WO-G001912.
XX
XX PR 07-DEC-1989; 89GB-0027722.
XX PR 07-DEC-1990; 90WO-GB01911.
XX
XX PA (BRB1-) BRIT BIO-TECHN LTD.
XX
XX PI Dawson KM, Edwards RM, Forman JM;
XX
XX WPI; 1991-208145/28.
XX DR N-PSDB; AAQ12546.
XX
XX Activatable fibrinolytic and antithrombic proteins - activated by
XX e.g. factor Xa, thrombin or activated protein C
XX
XX Claim 7; Fig 2 and Fig 4; 73pp; English.
XX
XX This protein is a plasminogen mutant cleavable by Factor Xa.
XX Activation is localised to the thrombus because cleavage to plasmin
XX is by an enzyme of the blood clotting pathway. Compositions
XX comprising the mutant plasminogen are used for treatment or
XX prevention of thrombosis, etc.
XX See AAQ12542-Q12558.
XX
XX SQ Sequence 813 AA;
XX
XX Query Match 66.1%; Score 39; DB 12; Length 813;
XX Best Local Similarity 46.2%; Pred. No. 92;
XX Matches 6; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
XX
QY 1 CGXPAIXPXXTGY 13
XX ||| : |||
Db 567 CGKPQVEPKKGY 579

RESULT 13
ABB61150
ID ABB61150 standard; Protein; 626 AA.
XX
XX AC ABB61150;
XX
XX DT 26-MAR-2002 (first entry)
XX
XX DE Drosophila melanogaster polypeptide SEQ ID NO 10242.
XX
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX OS Drosophila melanogaster.
XX
XX PN W0200171042-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 23-MAR-2001; 2001WO-US09231.
XX
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX
XX PA (PEXE ) PE CORP NY.
XX
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX DR N-PSDB; ABL05253.
XX
XX PT New isolated nucleic acid detection reagent for detecting 1003 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX PS Disclosure; SEQ ID NO 10242; 2lpp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 626 AA;
XX
XX Query Match 64.4%; Score 38; DB 22; Length 626;
XX Best Local Similarity 41.7%; Pred. No. 1.1e+02;
XX Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
XX
QY 1 CGXPAIXPXXTGY 12
XX ||| : |||
Db 406 CGTPTSLTPYSSG 417

RESULT 14
AAU54358
ID AAU54358 standard; Protein; 51 AA.
XX
XX AC AAU54358;
XX
XX DT 27-FEB-2002 (first entry)
XX
XX DE Propionibacterium acnes immunogenic protein #15254.
XX
XX SQ Sequence 51 AA;
XX
```

KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.

# Propionibacterium acnes.

WO200181581-A2.

01-NOV-2001.

20-APR-2001; 2001WO-US12865.

21-APR-2000; 2000US-199047P.

02-JUN-2000; 2000US-208841P.

07-JUL-2000; 2000US-216747P.

(CORI-) CORIXA CORP.

Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

L'maisonneuve J, Zhang Y, Jen S, Carter D;

WPI; 2001-616774/71.

N-PSDB; AAS9564.

Propionibacterium acnes polypeptides and nucleic acids useful for  
 vaccinating against and diagnosing infections, especially useful for  
 treating acne vulgaris -

Example 1; SEQ ID No 1553; 1069pp; English.

Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 polypeptides. The proteins and their associated DNA sequences are used in  
 the treatment, prevention and diagnosis of medical conditions caused by  
 P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 P. acnes is also involved in infections of bone, joints and the central  
 nervous system, however it is particularly involved in the inflammatory  
 lesions associated with acne vulgaris. A method for detecting the  
 presence or absence of P. acnes in a patient comprises contacting a  
 sample with a binding agent that binds to the proteins of the invention  
 and determining the amount of bound protein in the sample. The  
 polypeptides may be used as antigens in the production of antibodies  
 specific for P. acnes proteins. These antibodies can be used to  
 downregulate expression and activity of P. acnes polypeptides and  
 therefore treat P. acnes infections. The antibodies may also be used as  
 diagnostic agents for determining P. acnes presence, for example, by  
 enzyme linked immunosorbent assay (ELISA).

Note: The sequence data for this patent did not form part of the printed  
 specification, but was obtained in electronic format directly from WIPO  
 at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 51 AA;

Query Match 61.0%; Score 36; DB 22; Length 51;  
 Best Local Similarity 54.5%; Pred. No. 21;  
 Matches 6; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CGXPAXPXXT 11

Db 35 CGMPVIAPVST 45

RESULT 15

ABP09202

ID ABP09202 standard; Protein; 106 AA.

XX

AC ABP09202;

XX

DT 24-JUN-2002 (first entry)

XX

DE Human ORFX protein sequence SEQ ID NO:18386.

XX

Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
 hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
 degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 hypertension; hypothyroidism; cholesterol ester storage disease;  
 immune deficiency; immune disorder; infectious disease;  
 autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 myasthenia gravis.

Homo sapiens.

WO200192523-A2.

06-DEC-2001.

29-MAY-2001; 2001WO-US10836.

30-MAY-2000; 2000US-206132P.

29-AUG-2000; 2000US-228716P.

(CURA-) CURAGEN CORP.

Shinkets RA, Leach MD;

WPI; 2002-106308/14.

N-PSDB; ABN24954.

Novel human polypeptides and polynucleotides useful for diagnosing,  
 preventing and treating cardiovascular disease, neurodegenerative,  
 hyperproliferative disorders and autoimmune disorders -

Disclosure; SEQ ID 18386; 1037pp; English.

The present invention describes substantially purified human proteins  
 (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
 in the specification). ABN15762 to ABN27252 encode the human ORFX  
 proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
 treating or preventing a pathology associated with an ORFX-associated  
 disorder in humans, and in the manufacture of a medicament for treating a  
 syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
 sequences can be used in gene therapy. ORFX sequences can be used in the  
 treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
 osteoarthritis, neurodegenerative diseases, disorders related to organ  
 transplantation, cardiovascular diseases, diabetes mellitus, systemic  
 lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
 storage disease, various immune deficiencies and disorders, infectious  
 diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 bone degenerative disorders, or periodontal disease, and for gut  
 protection or regeneration and treatment of lung or liver fibrosis,  
 reperfusion injury in various tissues and conditions resulting from  
 systemic cytokine damage.

N.B. The sequence data for this patent did not form part of the printed  
 specification, but was obtained in electronic format directly from WIPO  
 at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 106 AA;

Query Match 61.0%; Score 36; DB 23; Length 106;

Best Local Similarity 54.5%; Pred. No. 43;

Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 GXPAXPXXTG 12

Db 7 GKPAVSPAHTG 17

Search completed: February 12, 2003, 10:22:23

Job time : 9.73134 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:17:50 ; Search time 3.2597 Seconds  
(without alignments)  
383.393 Million cell updates/sec

Title: US-10-036-371-3

Perfect score: 59

Sequence: 1 CGXPAIXPXXTGY 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 73:.\*  
1: PIR1:.\*  
2: PIR2:.\*  
3: PIR3:.\*  
4: PIR4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	88.1	244	2	S72219
2	52	88.1	263	2	S47537
3	43	72.9	1365	2	T30198
4	41	69.5	28	2	A61529
5	40	67.8	245	1	KVBOB
6	40	67.8	245	1	KYBOA
7	40	67.8	263	1	KYRTB
8	40	67.8	263	2	A11195
9	40	67.8	263	2	A11299
10	39	66.1	16	2	C61414
11	39	66.1	17	2	B61414
12	37	62.7	542	2	A82965
13	36	61.0	321	2	T07043
14	36	61.0	350	2	AD3170
15	36	61.0	373	2	AH3382
16	35	59.3	121	2	C71227
17	35	59.3	120	2	A71017
18	35	59.3	126	2	A23473
19	35	59.3	264	2	I38136
20	35	59.3	887	2	AG0535
21	35	59.3	4077	2	T17484
22	34	57.6	117	2	G75115
23	34	57.6	118	2	A24777
24	34	57.6	157	2	F84588
25	34	57.6	185	2	B82065
26	34	57.6	265	2	J50260
27	34	57.6	265	2	A38894
28	34	57.6	282	2	A84341
29	34	57.6	312	1	QQBE07

30	34	57.6	378	2	A33271
31	34	57.6	442	2	H71410
32	34	57.6	479	2	H90042
33	34	57.6	558	2	I56545
34	34	57.6	571	2	T52576
35	34	57.6	643	2	C90035
36	33	55.9	107	2	B72498
37	33	55.9	213	2	C89865
38	33	55.9	215	2	T22572
39	33	55.9	231	2	S37108
40	33	55.9	245	2	S06329
41	33	55.9	246	2	S00443
42	33	55.9	333	2	AD3284
43	33	55.9	569	2	A97648
44	33	55.9	569	2	AG2871
45	33	55.9	570	2	S42607

ALIGNMENTS

RESULT 1

S72219  
chymotrypsin B - Atlantic cod (fragments)  
C:Species: Gadus morhua (Atlantic cod)  
C:Date: 19-Mar-1998 #sequence\_revision 17-Apr-1998 #text\_change 07-Aug-1998  
C:Accession: S72219  
R:Leth-Larsen, R.; Asgeirsson, B.; Thorolfsson, M.; Norregaard-Madsen, M.; Hojrup, P.  
Biochim. Biophys. Acta 1297, 49-56, 1996  
A:Title: Structure of chymotrypsin variant B from Atlantic cod, Gadus morhua.  
A:Reference number: S72219; MUID:96439045; PMID:8841380  
A:Accession: S72219  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-14,15-244 <LET>  
C:Superfamily: trypsin; trypsin homology  
F:15-237/domain: trypsin homology <TRY>

Query Match 88.1%; Score 52; DB 2; Length 244;  
Best Local Similarity 69.2%; Pred. No. 0.011;  
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTGY 13

DB 1 CGSPAIOPOVTGY 13

RESULT 2

S47537  
chymotrypsin (EC 3.4.21.1) precursor - Atlantic cod  
C:Species: Gadus morhua (Atlantic cod)  
C:Date: 26-Dec-1994 #sequence\_revision 03-Aug-1995 #text\_change 22-Jun-1999  
R:Gundmundsdottir, A.; Oskarsson, S.; Eakin, A.E.; Craik, C.S.; Bjarnason, J.B.  
Biochim. Biophys. Acta 1219, 211-214, 1994  
A:Title: Atlantic cod cDNA encoding a psychrophilic chymotrypsinogen.  
A:Reference number: S47537; MUID:94368860; PMID:8086467  
A:Accession: S47537  
A:Molecule type: mRNA  
A:Residues: 1-263 <GUD>  
A:Cross-references: EMBL:X78490; NID:9468750; PIDN:CAAS5242.1; PID:9468751  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; protein digestion; serine proteinase  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-263/Product: chymotrypsin #status predicted <MAT>  
F:34-256/Domain: trypsin homology <TRY>  
F:75,120,213/Active site: His, Asp, Ser #status predicted

Query Match 88.1%; Score 52; DB 2; Length 263;  
Best Local Similarity 69.2%; Pred. No. 0.011;  
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTGY 13

```

Db      19  CGRPAISPVTGY 31
|||||
RESULT 3
T30198
alkaline phosphatase (EC 3.1.3.1) phoX, calcium-dependent - Volvox carteri
C:Species: Volvox carteri
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C:Accession: T30198
R:Hallmann, A. 274, 1691-1697, 1999
J. Biol. Chem. 274, 1691-1697, 1999
A:Title: Enzymes in the extracellular matrix of Volvox: an inducible, calcium-dependent
A:Reference number: 220772; PMID:99098915; PMID:9880549
A:Accession: T30198
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1365 <HML>
A:Cross-references: EMBL:AJ012458; NID:G4160585; PIDN:CAA10030.1; PID:G4160586
A:Experimental source: ssp. nagariensis, strain HK 10
C:Genetics:
A:Gene: phoX
C:Function:
A:Description: hydrolyzes phosphate esters
C:Superfamily: Volvox carteri alkaline phosphatase phoX
C:Keywords: magnesium; metalloprotein; phosphoric monoester hydrolase; zinc

Query Match      72.9%; Score 43; DB 2; Length 1365;
Best Local Similarity 53.8%; Pred. No. 3.2;
Matches 7; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy      1  CGXPAIXPXXXTG 13
|||||
Db      1063  CGTPATNPAPGY 1075

RESULT 4
A61529
chymotrypsin (EC 3.4.21.1) - Atlantic cod (fragments)
C:Species: Gadus morhua (Atlantic cod)
C:Date: 15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change 11-May-2000
C:Accession: A61529
R:Ageirsson, B.; Bjarnason, J.B.
Comp. Biochem. Physiol. B 99, 327-335, 1991
A:Title: Structural and kinetic properties of chymotrypsin from Atlantic cod (Gadus morhua)
A:Reference number: A61529; PMID:92111252; PMID:1764912
A:Accession: A61529
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-28 <ASG>
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; protein digestion; serine proteinase

Query Match      69.5%; Score 41; DB 2; Length 28;
Best Local Similarity 58.3%; Pred. No. 0.19;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy      1  CGXPAIXPXXXTG 12
|||||
Db      1  CGVPAIQPVLSG 12

RESULT 5
KYBOB
chymotrypsin (EC 3.4.21.1) B precursor - bovine
N:Alternate names: chymotrypsinogen B
C:Species: Bos primigenius taurus (cattle)
C:Date: 08-Oct-1981 #sequence_revision 08-Oct-1981 #text_change 18-Jul-1997
C:Accession: A00953
R:Smillie, L.B.; Furka, A.; Nagabhushan, N.; Stevenson, K.J.; Parkes, C.O.
Nature 218, 343-346, 1968
A:Title: Structure of chymotrypsinogen B compared with chymotrypsinogen A and trypsinogen
A:Reference number: A00953; PMID:68238908; PMID:5649671

```

---

```

A:Accession: A00953
A:Molecule type: protein
A:Residues: 1-245 <SMI>
C:Comment: Chymotrypsinogen B is synthesized, along with chymotrypsinogen A, in the acinar
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen
F:1-15/Domain: propetide #status experimental <PRO>
F:16-245/Product: chymotrypsin B #status experimental <MAT>
F:16-238/Domain: trypsin homology <TRY>
F:1-122,42-58,136-201,168-182,191-220/Disulfide bonds: #status experimental
F:57,102,195/Active site: His, Asp, Ser #status experimental

Query Match      67.8%; Score 40; DB 1; Length 245;
Best Local Similarity 58.3%; Pred. No. 2.4;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy      1  CGXPAIXPXXXTG 12
|||||
Db      1  CGVPAIQPVLSG 12

RESULT 6
KYBOA
chymotrypsin (EC 3.4.21.1) A precursor - bovine
N:Alternate names: chymotrypsinogen A
C:Species: Bos primigenius taurus (cattle)
C:Date: 07-May-1981 #sequence_revision 07-May-1981 #text_change 07-May-1999
C:Accession: A00235; A93158; S29650; A00952
R:Brown, J.R.; Hartley, B.S.
Biochem. J. 101, 214-228, 1966
A:Title: Location of disulphide bridges by diagonal paper electrophoresis. The disulphide
A:Reference number: A00235; PMID:67181721; PMID:5971783
A:Accession: A00235
A:Molecule type: protein
A:Residues: 1-101,'N',103-245 <BRO>
R:Blow, D.M.; Birktoft, J.J.; Hartley, B.S.
Nature 221, 337-340, 1969
A:Title: Role of a buried acid group in the mechanism of action of chymotrypsin.
A:Reference number: A93158; PMID:69106266; PMID:5764436
A:Content: annotation; revision to residue 102
R:Meloun, B.; Klueh, I.; Kostka, V.; Moravsek, L.; Prusik, Z.; Vanacek, J.; Keil, B.; Sorm,
Biochim. Biophys. Acta 130, 543-546, 1966
A:Title: Covalent structure of bovine chymotrypsinogen A.
A:Reference number: A90572; PMID:67183948; PMID:5972866
A:Accession: A93158
A:Molecule type: protein
A:Residues: 1-101,'N',103-245 <MEL>
A:Note: disulfide bonds were determined
R:Cutruzzola, F.; Ascenzi, P.; Barra, D.; Bolognesi, M.; Menegatti, E.; Sartl, P.; Schne
Biochim. Biophys. Acta 1161, 201-208, 1993
A:Title: Selective oxidation of Met-192 in bovine alpha-chymotrypsin. Effect on catalytic
A:Reference number: S29650; PMID:93160238; PMID:8431470
A:Accession: S29650
A:Molecule type: protein
A:Residues: 1-12,16-27,149-160,181-200 <CUT>
R:Smillie, L.B.; Hartley, B.S.
Biochem. J. 101, 232-241, 1966
A:Title: Histidine sequences in the active centres of some 'serine' proteinases.
A:Content: annotation; active site
A:Reference number: A90236; PMID:67181723; PMID:5971785
R:Birktoft, J.J.; Blow, D.M.; Henderson, R.; Steitz, T.A.
Philos. Trans. R. Soc. Lond. B257, 67-76, 1970
A:Title: The structure of alpha-chymotrypsin.
A:Reference number: A93754
A:Content: annotation; X-ray crystallography
C:Comment: Chymotrypsinogens are synthesized in the acinar cells of pancreas.
C:Superfamily: trypsin; further chymotryptic cleavage liberates the dipeptide Thr-147 and Asn-148 directly from chymotrypsinogen, which leads to the degraded form neochymotrypsin
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen
F:1-245/Product: chymotrypsinogen #status experimental <ZYM>

```

```

F:1-13,16-146,149-245/Product: alpha-chymotrypsin #status experimental <MPT>
F:16-238/Domain: trypsin homology <TRY>
F:1-122,42-58,136-201,168-182,191-220/Disulfide bonds: #status experimental
F:57,102,195/Active site: His, Asp, Ser #status experimental

Query Match          67.8%; Score 40; DB 1; Length 245;
Best Local Similarity 58.3%; Pred. No. 2.4;
Matches 7; Conservative 1; Mismatches 0; Gaps 0;

QY 1 CGXPAIXPXXTG 12
    |||||
Db 1 CGVPAIQPVLSG 12

RESULT 7
KYRTB
chymotrypsin (EC 3.4.21.1) B precursor - rat
N:Alternate names: chymotrypsinogen B
C:Species: Rattus norvegicus (Norway rat)
C>Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 18-Jun-1999
C:Accession: A22658
R:Bell, G.I.; Quinto, C.; Quiriga, M.; Valenzuela, P.; Craik, C.S.; Rutter, W.J.
J. Biol. Chem. 259, 14265-14270, 1984
A:Title: Isolation and sequence of a rat chymotrypsin B gene.
A:Reference number: A22658; MUID:85054881; PMID:6209274
A:Accession: A22658
A:Molecule type: DNA
A:Residues: 1-263 <BEL>
A:Cross-references: GB:K02298; NID:g203653; PIDN:AAA98732.1; PID:g203654
C:Genetics:
A:Introns: 18/1; 52/3; 79/2; 105/3; 166/1; 210/3
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase
F:1-18/Domain: signal sequence #status predicted <Sig>
F:19-33/Domain: propeptide #status predicted <PRO>
F:34-263/Product: chymotrypsin B #status predicted <MAT>
F:34-256/Domain: trypsin homology <TRY>
F:75,120,213/Active site: His, Asp, Ser #status predicted

Query Match          67.8%; Score 40; DB 1; Length 263;
Best Local Similarity 58.3%; Pred. No. 2.5;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 12
    |||||
Db 19 CGVPTIQPVLTG 30

RESULT 8
A21195
chymotrypsin (EC 3.4.21.1) 2 precursor - dog
C:Species: Canis lupus familiaris (dog)
C>Date: 27-Jul-1990 #sequence_revision 27-Jul-1990 #text_change 22-Jun-1999
C:Accession: A21195
R:Pinsky, S.D.; LaForge, K.S.; Luc, V.; Scheele, G.
Proc. Natl. Acad. Sci. U.S.A. 80, 7486-7490, 1983
A:Title: Identification of cDNA clones encoding secretory isoenzyme forms: sequence data
A:Reference number: A21195; MUID:84170253; PMID:6584866
A:Accession: A21195
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-263 <PIN>
A:Cross-references: GB:K01173; NID:g163945; PIDN:AAA30841.1; PID:g163946
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; protein digestion; serine proteinase
F:34-256/Domain: trypsin homology <TRY>
F:75,120,213/Active site: His, Asp, Ser #status predicted

Query Match          67.8%; Score 40; DB 2; Length 263;
Best Local Similarity 58.3%; Pred. No. 2.5;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 12
    |||||
Db 1 CGVPAIppVLSG 12

RESULT 11
B61414
chymotrypsin (EC 3.4.21.1) - painted turtle (fragment)
C:Species: Chrysemys picta (painted turtle)
C>Date: 19-Mar-1997 #sequence_revision 19-Dec-1997 #text_change 07-May-1999
C:Accession: B61414
R:Bhargava, A.K.; Barnard, E.A.
J. Mol. Evol. 2, 187-198, 1973
A:Title: Evolution in the pancreatic chymotrypsinogen series: N-terminal sequence deter
A:Reference number: A61414; MUID:76146602; PMID:4807189
A:Accession: B61414

```

```

Db 19 CGVPAIQPVLSG 30

RESULT 9
A31299
chymotrypsin (EC 3.4.21.1) precursor - human
C:Species: Homo sapiens (man)
C>Date: 08-Jun-1989 #sequence_revision 08-Jun-1989 #text_change 22-Jun-1999
C:Accession: A31299
R:Tomita, N.; Izumoto, Y.; Horii, A.; Doi, S.; Yokouchi, H.; Ogawa, M.; Mori, T.; Matsuba
Biochem. Biophys. Res. Commun. 158, 569-575, 1989
A:Title: Molecular cloning and nucleotide sequence of human pancreatic prechymotrypsinog
A:Reference number: A31299; MUID:89134264; PMID:2917002
A:Accession: A31299
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-263 <TOM>
A:Cross-references: GB:M24400; NID:g181189; PIDN:AAA52128.1; PID:g181190
C:Genetics:
A:Gene: GDB:CTRB1; CTRB
A:Cross-references: GDB:118820; OMIM:118890
A:Map position: 16q23.1-16q23.1
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; protein digestion; serine proteinase
F:34-256/Domain: trypsin homology <TRY>
F:75,120,213/Active site: His, Asp, Ser #status predicted

Query Match          67.8%; Score 40; DB 2; Length 263;
Best Local Similarity 58.3%; Pred. No. 2.5;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 12
    |||||
Db 19 CGVPAIHPVLSG 30

RESULT 10
C61414
chymotrypsin (EC 3.4.21.1) - slider turtle (fragment)
C:Species: Pseudemys scripta (slider)
C>Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 07-May-1999
C:Accession: C61414
R:Bhargava, A.K.; Barnard, E.A.
J. Mol. Evol. 2, 187-198, 1973
A:Title: Evolution in the pancreatic chymotrypsinogen series: N-terminal sequence deter
A:Reference number: A61414; MUID:76146602; PMID:4807189
A:Accession: C61414
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-16 <BHA>
C:Keywords: hydrolase; protein digestion; serine proteinase

Query Match          66.1%; Score 39; DB 2; Length 16;
Best Local Similarity 58.3%; Pred. No. 0.27;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 12
    |||||
Db 1 CGVPAIppVLSG 12

RESULT 11
B61414
chymotrypsin (EC 3.4.21.1) - painted turtle (fragment)
C:Species: Chrysemys picta (painted turtle)
C>Date: 19-Mar-1997 #sequence_revision 19-Dec-1997 #text_change 07-May-1999
C:Accession: B61414
R:Bhargava, A.K.; Barnard, E.A.
J. Mol. Evol. 2, 187-198, 1973
A:Title: Evolution in the pancreatic chymotrypsinogen series: N-terminal sequence deter
A:Reference number: A61414; MUID:76146602; PMID:4807189
A:Accession: B61414

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:04:45 ; Search time 1.74627 Seconds  
(without alignments)  
308.768 Million cell updates/sec

Title: US-10-036-371-3  
Perfect score: 59  
Sequence: 1 CGXPAIXXTXY 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	88.1	245	1 CTRB_GADMO	P80646 Gadus morhu
2	52	88.1	243	1 CTRB_GADMO	P47796 Gadus morhu
3	40	67.8	245	1 CTRB_BOVIN	P00766 bos taurus
4	40	67.8	245	1 CTRB_BOVIN	P00767 bos taurus
5	40	67.8	263	1 CTR2_CANFA	P04813 canis famil
6	40	67.8	263	1 CTRB_HUMAN	P17538 homo sapien
7	40	67.8	263	1 CTRB_RAT	P07338 rattus norv
8	38	64.4	591	1 HE_HEMPU	P91953 hemicentrot
9	38	64.4	626	1 DALY_DROME	Q24114 drosophila
10	35	59.3	264	1 CTRB_HUMAN	P40313 homo sapien
11	35	59.3	948	1 CTRB_MOUSE	Q92082 mus musculu
12	35	59.3	955	1 CTRD_HUMAN	Q92082 mus musculu
13	34	57.6	265	1 SERI_DROME	P17205 drosophila
14	34	57.6	282	1 FPNK_HALNI	Q9hm77 halobacteri
15	34	57.6	312	1 US10_HSV11	P06486 herpes simp
16	34	57.6	378	1 PEM1_PHACH	Q02567 phanerocae
17	34	57.6	558	1 GPC1_RAT	P35053 rattus norv
18	33	55.9	231	1 CTR2_CAEBL	P34682 caenorhabdi
19	33	55.9	245	1 CB11_LYCF5	P12360 lycopersico
20	33	55.9	570	1 UR11_RHIME	P42885 rhizobium m
21	33	55.9	730	1 DPOL_HPBV4	P12993 hepatitis b
22	33	55.9	750	1 DPOL_HPBV2	P03155 hepatitis b
23	33	55.9	763	1 DPOL_HPBV3	Q02314 hepatitis b
24	33	55.9	832	1 DPOL_HPBVA	P24024 hepatitis b
25	33	55.9	832	1 DPOL_HPBVL	P12900 hepatitis b
26	33	55.9	832	1 DPOL_HPBVY	P04484 hepatitis b
27	33	55.9	832	1 DPOM_HPBVY	P18700 hepatitis b
28	33	55.9	842	1 DPOL_HPBVM	P17393 hepatitis b
29	33	55.9	843	1 DPOL_HPBVJ	P17394 hepatitis b
30	33	55.9	843	1 DPOL_HPBVO	P17395 hepatitis b
31	33	55.9	843	1 DPOL_HPBVR	P03157 hepatitis b
32	33	55.9	843	1 DPOL_HPBVS	Q05486 hepatitis b
33	33	55.9	843	1 DPOL_HPBVT	Q05486 hepatitis b

34	33	55.9	845	1 DPOL_HPBV2	P03159 hepatitis b
35	33	55.9	845	1 DPOL_HPBV9	P17100 hepatitis b
36	33	55.9	1329	1 FTSK_ECOLI	P46889 escherichia
37	33	55.9	2871	1 FBN1_BOVIN	P98133 bos taurus
38	33	55.9	2871	1 FBN1_PIG	Q9tv36 sus scrofa
39	32	54.2	127	1 TFF2_PIG	P01359 sus scrofa
40	32	54.2	129	1 TFF2_HUMAN	Q03403 homo sapien
41	32	54.2	129	1 TFF2_MOUSE	Q03404 mus musculu
42	32	54.2	129	1 TFF2_RAT	Q09030 rattus norv
43	32	54.2	322	1 YPX4_CAEBL	Q20257 caenorhabdi
44	32	54.2	344	1 SNXG_RAT	P57769 rattus norv
45	32	54.2	510	1 CP5Q_CANNA	Q12587 candida mal

ALIGNMENTS

RESULT 1				
CTRB_GADMO	CTRB_GADMO	STANDARD;	PRT;	245 AA.
ID	P80646;			
AC	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Chymotrypsin B (EC 3.4.21.1).			
OS	Gadus morhua (Atlantic cod).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;			
OC	Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.			
OX	NCBI_TaxID=8049;			
RN	[1]			
RN	[2]			
RP	SEQUENCE.			
RC	TISSUE=Pyloric caeca;			
RX	MEDLINE=36439045; PubMed=8841380;			
RA	Leth-Larsen R., Asgeirsson B., Thorolfsson M., Noerregaard-Madsen M.,			
RA	Hoejrup P.;			
RT	"Structure of chymotrypsin variant B from Atlantic cod, Gadus			
RT	morhua.";			
RL	Biochim. Biophys. Acta 1257:49-56 (1996).			
RN	[2]			
RP	SEQUENCE OF 1-12 AND 16-31.			
RC	TISSUE=Pyloric caeca;			
RX	MEDLINE=92111252; PubMed=1764912;			
RA	Asgeirsson B., Bjarnason J.B.;			
RT	"Structural and kinetic properties of chymotrypsin from Atlantic cod			
RT	(Gadus morhua). Comparison with bovine chymotrypsin.";			
RL	Comp. Biochem. Physiol. 99B:327-335 (1991).			
CC	-!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr- -Xaa, Trp- -Xaa,			
CC	Phe- -Xaa, Leu- -Xaa.			
CC	-!- SUBCELLULAR LOCATION: Extracellular.			
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.			
DR	HSP; P00766; 1CHG.			
DR	WEROPS; S01152; 1.			
DR	InterPro; IPR001314; Chymotrypsin.			
DR	InterPro; IPR001254; Ser_protease_Try.			
DR	Pfam; PF00089; trypsin; 1.			
DR	PRINTS; PR00722; CHYMOTRYPSIN.			
DR	SMART; SM00020; Tryp SPC; 1.			
DR	PROSITE; PS00240; TRYPSIN_DOM; 1.			
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.			
DR	PROSITE; PS00135; TRYPSIN_SER; 1.			
KW	Hydrolase; Serine protease; Digestion; Pancreas; Zymogen.			
FT	CHAIN 1 13			
FT	ACT_SITE 16 245			
FT	ACT_SITE 57 57			
FT	ACT_SITE 101 101			
FT	ACT_SITE 195 195			
FT	DISULFID 1 121			
FT	DISULFID 42 58			
FT	DISULFID 135 201			
FT	DISULFID 167 182			
FT	DISULFID 191 220			
FT	CONFLICT 9 11			
FT	QVT -> VIS (IN REF. 2).			

```

FT CONFLICT 26 26 S -> T (IN REF. 2).
FT CONFLICT 28 29 PW -> Y (IN REF. 2).
SQ SEQUENCE 245 AA; 26260 MW; 74FE0D425517AB02 CRC64;

Query Match
Best Local Similarity 88.1%; Score 52; DB 1; Length 245;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CGXPAIXPXXTGY 13
Db 1 CGSPAIPQVTGY 13

RESULT 2
CTRA_GADMO
ID CTRA_GADMO STANDARD; PRT; 263 AA.
AC P47756;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymotrypsin A precursor (EC 3.4.21.1).
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.
OX NCBI_TaxID=8049;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pyloric caeca;
RX MEDLINE=9436860; PubMed=8086467;
RA Gudmundsdottir A., Oskarsson S., Eakin A.E., Craik C.S.,
RA Bjarnason J.B.;
RT "Atlantic cod cDNA encoding a psychrophilic chymotrypsinogen.";
RL Biochim. Biophys. Acta 1219:211-214(1994).
RN [2]
RP SEQUENCE OF 19-30 AND 34-49.
RC TISSUE=Pyloric caeca;
RX MEDLINE=9211252; PubMed=1764912;
RA Ageelsson B., Bjarnason J.B.;
RT "Structural and kinetic properties of chymotrypsin from Atlantic cod (Gadus morhua). Comparison with bovine chymotrypsin.";
RL Comp. Biochem. Physiol. 99B:327-335(1991).
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
CC Phe-|-Xaa, Leu-|-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; X78490; CAA55242.1; -.
DR HSP6; P00766; ICHG.
DR MEROPS; S01.152; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal.
FT SIGNAL 1 18
FT CHAIN 19 263 CHYMOTRYPSIN A.
FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 19 140 BY SIMILARITY.

```

```

FT DISULFID 60 76 BY SIMILARITY.
FT DISULFID 154 219 BY SIMILARITY.
FT DISULFID 186 200 BY SIMILARITY.
FT DISULFID 209 238 BY SIMILARITY.
FT CONFLICT 21 21 R -> S (IN REF. 2).
FT CONFLICT 25 25 S -> Q (IN REF. 2).
FT CONFLICT 29 29 T -> S (IN REF. 2).
FT CONFLICT 44 44 S -> T (IN REF. 2).
FT CONFLICT 46 46 S -> Y (IN REF. 2).
SQ SEQUENCE 263 AA; 28294 MW; 47AAC699A0A64FBB CRC64;

Query Match
Best Local Similarity 88.1%; Score 52; DB 1; Length 263;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CGXPAIXPXXTGY 13
Db 19 CGSPAIPQVTGY 31

RESULT 3
CTRA_BOVIN
ID CTRA_BOVIN STANDARD; PRT; 245 AA.
AC P00766;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymotrypsinogen A (EC 3.4.21.1).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE, DISULFIDE BONDS, AND ACTIVE SITE.
RX MEDLINE=67181721; PubMed=5971783;
RA Brown J.R., Hartley B.S.;
RT "Location of disulphide bridges by diagonal paper electrophoresis. The disulphide bridges of bovine chymotrypsinogen A.";
RL Biochem. J. 101:214-228(1966).
RN [2]
RP REVISION TO 102.
RX MEDLINE=6910266; PubMed=5764436;
RA Blow D.M., Birkoft J.J., Hartley B.S.;
RT "Role of a buried acid group in the mechanism of action of chymotrypsin.";
RL Nature 221:337-340(1969).
RN [3]
RP PRELIMINARY SEQUENCE.
RA Hartley B.S.;
RT "Amino-acid sequence of bovine chymotrypsinogen-A.";
RL Nature 201:1284-1287(1964).
RN [4]
RP SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE=67183948; PubMed=5972866;
RA Meloun B., Klueh I., Kostka V., Moravec L., Prusik Z., Vanacek J.,
RA Keil B., Sorm P.;
RT "Covalent structure of bovine chymotrypsinogen A.";
RL Biochim. Biophys. Acta 130:543-546(1966).
RN [5]
RP ACTIVE SITE.
RX MEDLINE=67181723; PubMed=5971785;
RA Smillie L.B., Hartley B.S.;
RT "Histidine sequences in the active centres of some 'serine' proteinases."
RL Biochem. J. 101:232-241(1966).
RN [6]
RP X-RAY CRYSTALLOGRAPHY.
RX MEDLINE=72035052; PubMed=4399050;
RA Birkoft J.J., Blow D.M., Henderson R., Steitz T.A.;
RT "I. Serine proteinases. The structure of alpha-chymotrypsin.";
RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 257:67-76(1970).
RN [7]

```



RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF CHYMOTRYPSINOGEN.  
RA MEDLINE=70177557; PubMed=5442169;  
RA Freer S.T., Kraut J., Robertus J.D., Wright H.T., Xuong N.H.;  
RT "Chymotrypsinogen: 2.5-A crystal structure, comparison with alpha-  
RT chymotrypsin, and implications for zymogen activation.";  
RL Biochemistry 9:1997-2009(1970).  
RN [6].  
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF GAMMA-CHYMOTRYPSIN.  
RA MEDLINE=82078042; PubMed=6914398;  
RA Cohen G.H., Silvertown E.W., Davies D.R.;  
RT "Refined crystal structure of gamma-chymotrypsin at 1.9-A resolution.  
RT Comparison with other pancreatic serine proteases.";  
RL J. Mol. Biol. 148:449-479(1981).  
RN [9].  
RP X-RAY CRYSTALLOGRAPHY (1.68 ANGSTROMS) OF ALPHA-CHYMOTRYPSIN.  
RA MEDLINE=86011575; PubMed=4046030;  
RA Teukada H., Blow D.M.;  
RT "Structure of alpha-chymotrypsin refined at 1.68-A resolution.";  
RL J. Mol. Biol. 184:703-711(1985).  
CC -|- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,  
CC Phe-|-Xaa, Leu-|-Xaa.  
CC -|- SUBCELLULAR LOCATION: Extracellular.  
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
CC -|- DATABASE: NAME=Worthington enzyme manual;  
CC WWW="http://www.worthington-biochem.com/manual/C/CHY.html".  
DR PIR; A00952; KYBOA.  
DR PDB; 2CGA; 15-APR-90.  
DR PDB; 2CHA; 31-MAY-84.  
DR PDB; 4CHA; 29-OCT-85.  
DR PDB; 5CHA; 16-OCT-87.  
DR PDB; 6CHA; 16-OCT-87.  
DR PDB; 1CHG; 27-JAN-84.  
DR PDB; 1CHO; 16-JUL-88.  
DR PDB; 2GCH; 31-MAY-84.  
DR PDB; 3GCH; 15-OCT-92.  
DR PDB; 4GCH; 15-OCT-90.  
DR PDB; 5GCH; 15-OCT-90.  
DR PDB; 6GCH; 15-OCT-90.  
DR PDB; 7GCH; 15-OCT-90.  
DR PDB; 8GCH; 15-JUL-93.  
DR PDB; 1GCT; 15-OCT-91.  
DR PDB; 2GCT; 15-OCT-91.  
DR PDB; 3GCT; 15-OCT-91.  
DR PDB; 1ACB; 31-OCT-93.  
DR PDB; 1GMC; 31-OCT-93.  
DR PDB; 1GMD; 31-OCT-93.  
DR PDB; 1CGI; 30-APR-94.  
DR PDB; 1GGJ; 30-APR-94.  
DR PDB; 1GCD; 22-JUN-94.  
DR PDB; 1GHA; 22-JUN-94.  
DR PDB; 1GHB; 22-JUN-94.  
DR PDB; 1GMH; 30-SEP-94.  
DR PDB; 2GMT; 01-NOV-94.  
DR PDB; 1MTN; 17-AUG-96.  
DR PDB; 1AB9; 20-AUG-97.  
DR PDB; 1AFQ; 17-SEP-97.  
DR PDB; 1CA0; 23-JUL-97.  
DR PDB; 1CBW; 23-JUL-97.  
DR PDB; 1VGC; 12-NOV-97.  
DR PDB; 2VGC; 12-NOV-97.  
DR PDB; 3VGC; 12-NOV-97.  
DR PDB; 4VGC; 12-NOV-97.  
DR PDB; 1HJA; 14-JAN-98.  
DR MEROPS; S01.001; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser. protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
Hydrolase; Serine protease; Digestion; Pancreas; Zymogen;

KW 3D-structure. 1 13 CHYMOTRYPSIN A, A CHAIN.  
FT CHAIN 16 146 CHYMOTRYPSIN A, B CHAIN.  
FT CHAIN 149 245 CHYMOTRYPSIN A, C CHAIN.  
FT ACT\_SITE 57 57 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 102 102 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 195 195 CHARGE RELAY SYSTEM.  
FT DISULFID 1 122  
FT DISULFID 42 58  
FT DISULFID 136 201  
FT DISULFID 168 182  
FT DISULFID 191 220  
FT HELIX 12 15  
FT STRAND 20 21  
FT TURN 24 25  
FT TURN 28 29  
FT STRAND 30 34  
FT STRAND 40 46  
FT STRAND 51 54  
FT TURN 57 58  
FT TURN 62 63  
FT STRAND 65 68  
FT TURN 69 69  
FT STRAND 72 72  
FT TURN 73 74  
FT STRAND 81 90  
FT TURN 92 93  
FT TURN 95 95  
FT TURN 96 99  
FT STRAND 100 100  
FT TURN 101 101  
FT STRAND 104 108  
FT STRAND 122 122  
FT TURN 126 127  
FT TURN 132 133  
FT STRAND 135 140  
FT TURN 146 148  
FT STRAND 154 154  
FT STRAND 156 163  
FT HELIX 165 168  
FT TURN 169 172  
FT HELIX 173 175  
FT STRAND 180 184  
FT TURN 182 193  
FT TURN 185 196  
FT STRAND 198 203  
FT TURN 204 205  
FT STRAND 206 216  
FT TURN 218 219  
FT TURN 222 223  
FT STRAND 225 230  
FT HELIX 231 233  
FT TURN 234 234  
FT HELIX 235 243  
SQ SEQUENCE 245 AA; 25666 MW; 91A9F28E2F3E3142 CRC64;  
Query Match 67.8%; Score 40; DB 1; Length 245;  
Best Local Similarity 58.3%; Pred. No. 0.53;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 1 CGXPALXPXXTG 12  
Db 1 CGVPAIQVLSG 12  
RESULT 4  
ID\_CTRB\_BOVIN STANDARD; PRT; 245 AA.  
AC P00767;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Chymotrypsinogen B (EC 3.4.21.1).

OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE, DISULFIDE BONDS, AND ACTIVE SITE.  
 RX MEDLINE=68238908; PubMed=5649671;  
 RA Smillie L.B., Furka A., Nagabhushan N., Stevenson K.J., Parkes C.O.;  
 RT "Structure of chymotrypsinogen B compared with chymotrypsinogen A and  
 trypsinogen.";  
 RL Nature 218:343-346(1968).  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr|-Xaa, Trp|-Xaa,  
 Phe|-Xaa, Leu|-Xaa.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -1- DATABASE: NAME=worthington enzyme manual;  
 WWW="http://www.worthington-biochem.com/manual/c/CHY.html".  
 CC PIR; A00953; KYBOB.  
 DR HSP; P00766; IACB.  
 DR MEROPS; S01.152;  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen.  
 FT CHAIN 1 13 CHYMOTRYPSIN B, A CHAIN.  
 FT CHAIN 16 146 CHYMOTRYPSIN B, B CHAIN.  
 FT CHAIN 149 245 CHYMOTRYPSIN B, C CHAIN.  
 FT ACT\_SITE 57 57 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 102 102 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 195 195 CHARGE RELAY SYSTEM.  
 FT DISULFID 1 122 CHARGE RELAY SYSTEM.  
 FT DISULFID 42 58  
 FT DISULFID 136 201  
 FT DISULFID 168 182  
 FT DISULFID 191 220  
 SQ SEQUENCE 245 AA; 25755 MW; 678016446FF5FEB5 CRC64;  
 Query Match 67.8%; Score 40; DB 1; Length 245;  
 Best Local Similarity 58.3%; Pred. No. 0.53;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 CGXPAIXPXXTG 12  
 Db ||||| :  
 1 CGVPAIQPVLSG 12  
 RESULT 5  
 ID\_CTR2\_CANFA STANDARD; PRT; 263 AA.  
 AC P04813;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Chymotrypsinogen 2 precursor (EC 3.4.21.1).  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84170253; PubMed=6584866;  
 RA Pinsky S.D., Laforge K.S., Luc V., Scheele G.;  
 RT "Identification of cDNA clones encoding secretory isoenzyme forms:  
 sequence determination of canine pancreatic precymotrypsinogen 2  
 mRNA.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:7486-7490(1983).  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr|-Xaa, Trp|-Xaa,

Phe|-Xaa, Leu|-Xaa.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; K01173; AAA30841.1; -.  
 DR PIR; A21195; A21195.  
 DR HSP; P00766; IACB.  
 DR MEROPS; S01.152;  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 263 CHYMOTRYPSINOGEN 2.  
 FT CHAIN 19 31 CHYMOTRYPSIN 2, A CHAIN.  
 FT CHAIN 34 164 CHYMOTRYPSIN 2, B CHAIN.  
 FT CHAIN 167 263 CHYMOTRYPSIN 2, C CHAIN.  
 FT ACT\_SITE 75 75 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 120 120 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 213 213 CHARGE RELAY SYSTEM.  
 FT DISULFID 19 140 BY SIMILARITY.  
 FT DISULFID 60 76 BY SIMILARITY.  
 FT DISULFID 154 219 BY SIMILARITY.  
 FT DISULFID 186 200 BY SIMILARITY.  
 FT DISULFID 209 238 BY SIMILARITY.  
 SQ SEQUENCE 263 AA; 27787 MW; 2A2F49D813B3961 CRC64;  
 Query Match 67.8%; Score 40; DB 1; Length 263;  
 Best Local Similarity 58.3%; Pred. No. 0.56;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 CGXPAIXPXXTG 12  
 Db ||||| :  
 19 CGVPAIQPVLSG 30  
 RESULT 6  
 ID\_CTRB\_HUMAN STANDARD; PRT; 263 AA.  
 AC P17538;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Chymotrypsinogen B precursor (EC 3.4.21.1).  
 GN CTRB1 OR CTRB (Human).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=89134284; PubMed=2917002;  
 RA Tomita N., Izumoto Y., Horii A., Yokouchi H., Ogawa M.,  
 RA Mori T., Matsubara K.;  
 RT "Molecular cloning and nucleotide sequence of human pancreatic  
 precymotrypsinogen cDNA.";  
 RL Biochem. Biophys. Res. Commun. 158:569-575(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.

```

RC TISSUE-Pancreas;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
CC Phe-|-Xaa, Leu-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M24400; AA52128.1; -.
CC EMBL: BC005385; AA00385.1; -.
CC PIR: A31299; A31299.
CC HSP: P00766; ICHG.
CC MEROPS: S01.152; -.
CC Genew: HGNC:2521; CTRB1.
CC MIM: 118890; -.
CC InterPro: IPR001314; Chymotrypsin.
CC InterPro: IPR001254; Ser_protease_Try.
CC Pfam: PF00089; trypsin; 1.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC SMART: SM00020; Tryp_SPC; 1.
CC PROSITE: PS0240; TRYPSIN_DOM; 1.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal.
CC
CC FT SIGNAL 1 18 CHYMOTRYPSINOGEN B.
CC FT CHAIN 19 263 CHYMOTRYPSIN B, A CHAIN.
CC FT CHAIN 34 164 CHYMOTRYPSIN B, B CHAIN.
CC FT CHAIN 167 263 CHYMOTRYPSIN B, C CHAIN.
CC FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT DISULFID 19 140 BY SIMILARITY.
CC FT DISULFID 60 76 BY SIMILARITY.
CC FT DISULFID 154 219 BY SIMILARITY.
CC FT DISULFID 186 200 BY SIMILARITY.
CC FT DISULFID 209 238 BY SIMILARITY.
CC SQ SEQUENCE 263 AA; 27870 MW; 4C1C055A490B8701 CRC64;

Query Match 67.8%; Score 40; DB 1; Length 263;
Best Local Similarity 58.3%; Pred. No. 0.56;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 12
DB 19 CGVPAIHVLSG 30

RESULT 7
CTRB RAT STANDARD; PRT; 263 AA.
AC P07338;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymotrypsinogen B precursor (EC 3.4.21.1).
GN CTRB1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85054881; PubMed=6209274;
RA Bell G.I., Quinto C., Quiroga M., Valenzuela P., Craik C.S.,

```

```

RA Rutter W.J.;
RL "Isolation and sequence of a rat chymotrypsin B gene.";
RL J. Biol. Chem. 259:14265-14270(1984).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
CC Phe-|-Xaa, Leu-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: K02298; AAA98732.1; -.
CC PIR: A22658; KYTB.
CC HSP: P00766; ICHG.
CC MEROPS: S01.152; -.
CC InterPro: IPR001314; Chymotrypsin.
CC InterPro: IPR001254; Ser_protease_Try.
CC Pfam: PF00089; trypsin; 1.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC SMART: SM00020; Tryp_SPC; 1.
CC PROSITE: PS0240; TRYPSIN_DOM; 1.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal.
CC
CC FT SIGNAL 1 18 CHYMOTRYPSINOGEN B.
CC FT CHAIN 19 263 CHYMOTRYPSIN B, A CHAIN.
CC FT CHAIN 34 164 CHYMOTRYPSIN B, B CHAIN.
CC FT CHAIN 167 263 CHYMOTRYPSIN B, C CHAIN.
CC FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT DISULFID 19 140 BY SIMILARITY.
CC FT DISULFID 60 76 BY SIMILARITY.
CC FT DISULFID 154 219 BY SIMILARITY.
CC FT DISULFID 186 200 BY SIMILARITY.
CC FT DISULFID 209 238 BY SIMILARITY.
CC SQ SEQUENCE 263 AA; 27849 MW; ACAFDACF8C4DA6D CRC64;

Query Match 67.8%; Score 40; DB 1; Length 263;
Best Local Similarity 58.3%; Pred. No. 0.56;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 12
DB 19 CGVTPIQVLTG 30

RESULT 8
HE HEMPU STANDARD; PRT; 591 AA.
ID HE HEMPU
AC P91953;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hatching enzyme precursor (EC 3.4.24.12) (HE) (HEZ) (Envelysin)
DE (Sea-urchin-hatching proteinase).
OS Hemichordatus pulcherrimus (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
OC Hemichordatus.
OX NCBI_TaxID=7650;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 170-220 AND 504-528.
RC TISSUE=Blascula;
RX MEDLINE=97332490; PubMed=9188724;
RA Nomura K., Shimizu T., Kinoh H., Sendai Y., Inomata M., Suzuki N.;
RA "Sea urchin hatching enzyme (envelysin) (envelysin): cDNA cloning and deprivation

```

of protein substrate specificity by autolytic degradation.";

RT Biochemistry 36:7225-7238(1997).

[2]

RP CLEAVAGE SPECIFICITY.

RA MEDLINE=91283448; PubMed=1711895;

RA Nomura K., Tanaka H., Kikawa Y., Yamaguchi M., Suzuki N.;

RT "The specificity of sea urchin hatching enzyme (envelysin) places it

in the mammalian matrix metalloproteinase family.";

RT Biochemistry 30:6115-6123(1991).

[3]

RP STEREO-SPECIFICITY.

RP MEDLINE=93223852; PubMed=8467915;

RA Nomura K., Suzuki N.;

RT "Stereo-specific inhibition of sea urchin envelysin (hatching enzyme)

by a synthetic autoinhibitor peptide with a cysteine-switch consensus

sequence.";

RT FEBS Lett. 321:84-88(1993).

CC -!- FUNCTION: ALLOWS THE SEA URCHIN TO DIGEST THE PROTECTIVE ENVELOPE

DERIVED FROM THE EGG EXTRACELLULAR MATRIX THUS ALLOWING THE SEA

URCHIN TO SWIM FREELY.

CC -!- CATALYTIC ACTIVITY: Preferential cleavage: on the amino side of

bulky hydrophobic residues, -Leu, -Ile, -Phe, as well as -Tyr.

CC -!- SUBUNIT: DURING HATCHING, THE 50 kDa MATURE ENZYME IS

AUTOLYTICALLY CLEAVED TO PRODUCE A MAJOR 38 kDa AND A MINOR 15 kDa

FORM WHICH MAY BE DISULFIDE LINKED. SUBSEQUENT CLEAVAGE OF THE 38

kDa SPECIES YIELDS A 32 kDa NON-SPECIFIC PROTEASE.

CC -!- DEVELOPMENTAL STAGE: EMBRYO, BLASTULA STAGE. HIGHEST ACTIVITY AT

12.5 HRS EMBRYO STAGE.

CC -!- DOMAIN: THERE ARE TWO DISTINCT DOMAINS IN THIS PROTEIN; THE

CATALYTIC N-TERMINAL, AND THE C-TERMINAL WHICH IS INVOLVED IN

SUBSTRATE SPECIFICITY.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A.

CC -!- SIMILARITY: CONTAINS 1 HEMOPEXIN-LIKE DOMAIN.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR ENBL; AB000719; BAA19171.1; -.

DR HSSP; P03956; ICGI.

DR MEROPS; M10.010; -.

DR InterPro; IPR000585; Hemoexin.

DR InterPro; IPR001818; Matrixin.

DR InterPro; IPR00130; Zn\_Mtpeptidse.

DR Pfam; PF00045; hemoexin; 4.

DR Pfam; PF00413; Peptidase\_M10; 1.

DR PRINTS; PR00138; MATRIXIN.

DR SMART; SM00120; HX; 4.

DR SMART; SM00235; ZnMc; 1.

DR PROSITE; PS00024; HEMOPEXIN; 1.

DR PROSITE; PS00142; ZINC\_PROTEASE; 1.

DR PROSITE; PS00546; CYSTEINE\_SWITCH; 1.

DR Hydrolase; Metalloprotease; zinc; Glycoprotein; Zymogen;

KW Autocatalytic cleavage; Signal; Calcium.

FT SIGNAL 1 18

FT PROPEP 19 169

FT CHAIN 170 591

FT CHAIN 170 450

FT CHAIN 170 503

FT CHAIN 504 591

FT CHAIN 62 76

FT DOMAIN 336 380

FT DOMAIN 380 589

FT DOMAIN 448 451

FT SITE 162 162

FT METAL 286 286

FT ACT\_SITE 287 287

FT METAL 290 290

FT SITE 450 451

CLEAVAGE (AUTOLYTIC DURING HATCHING).

FT SITE 504

FT DISULFID 383

FT CARBOHYD 129

FT CARBOHYD 144

FT CARBOHYD 578

FT CARBOHYD 588

FT CONFLICT 195

FT CONFLICT 197

FT CONFLICT 203

FT CONFLICT 215

FT CONFLICT 220

FT CONFLICT 509

FT CONFLICT 511

FT CONFLICT 518

FT CONFLICT 522

SO SEQUENCE 591 AA; 66126 MW; 5DCB48C6758C70D CRC64;

Query Match 64.4%; Score 38; DB 1; Length 591;

Best Local Similarity 58.3%; Pred. No. 3.2;

Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CGXPALXPXTG 12

Db 162 CGVPDILPYVTG 173

-----

RESULT 9

DALY\_DROME STANDARD; PRT; 626 AA.

ID DALY\_DROME

AC Q24114; Q9VQS8;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Division abnormally delayed protein precursor (Dally protein).

GN DALLY OR CG4974.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96102814; PubMed=8582281;

RA Nakato H., Futch T.A., Selleck S.B.;

RT "The division abnormally delayed (dally) gene: a putative integral

membrane proteoglycan required for cell division patterning during

postembryonic development of the nervous system in Drosophila.";

RL Development 121:3687-3702(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkeley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Cocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.P.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Beeson R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Bellow K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabriellia A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Ruskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasaman D.A., Weinstein G.M., Weisenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of Drosophila melanogaster";  
PL Science 287:2185-2195(2000).  
CC -!- FUNCTION: HEPARAN SULFATE CELL SURFACE PROTEOGLYCAN. REQUIRED FOR  
CC CELL DIVISION PATTERNING DURING POST-EMBRYONIC DEVELOPMENT OF THE  
CC NERVOUS SYSTEM.  
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
CC (potential).  
CC -!- SIMILARITY: BELONGS TO THE GLYPICAN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; X71874; CAAS0710.1; -;  
CC EMBL; X71877; CAAS0711.1; -;  
CC HSP; P00763; LDPO.  
CC MEROPS; S01.256; -;  
CC Genew; HGNC:2524; CTRL.  
CC MW; 118888; -;  
CC InterPro; IPR001314; Chymotrypsin.  
CC InterPro; IPR001254; Ser. protease\_Try.  
CC Pfam; PF00089; trypsin; 1.  
CC PRINTS; PR00722; CHYMOTRYPSIN.  
CC SMART; SM00020; Tryp\_Spc; 1.  
CC PROSITE; PS00240; TRYPSIN\_DOM; 1.  
CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
CC Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal.  
KW SIGNAL 1 18  
FT PROPEP 19 33 ACTIVATION PEPTIDE (POTENTIAL).  
FT CHAIN 34 264 CHYMOTRYPSIN-LIKE PROTEASE CTRL-1.  
FT ACT\_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 121 121 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 214 214 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT CARBOHYD 114 114 N-LINKED (GLCNAC...) (POTENTIAL).  
FT DISULFID 19 141 BY SIMILARITY.  
FT DISULFID 60 76 BY SIMILARITY.  
FT DISULFID 155 220 BY SIMILARITY.  
FT DISULFID 187 201 BY SIMILARITY.  
FT DISULFID 210 239 BY SIMILARITY.  
SQ SEQUENCE 264 AA; 28002 MW; 3F629F02FA6DDFB4 CRC64;  
  
Query Match 59.3%; Score 35; DB 1; Length 264;  
Best Local Similarity 75.0%; Pred. No. 6;  
Matches 6; Conservative 0; Mismatches 0; Indels 2; Gaps 0;  
  
QY 1 CGXPAIXP 8  
Db 19 CGIPAIXP 26  
|||:|:|:  
ID -CHRD MOUSE STANDARD; PRT; 948 AA.  
RESULT 11  
CHRD MOUSE  
AC Q320E2;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Chordin precursor.  
GN CHRD.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B6SJL/F1;  
RA Lu B., Bachiller D., Agius E., Piccolo S., De Robertis E.M.;  
RT "BMP-binding domains in the chordin secreted protein.";  
  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Ruskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasaman D.A., Weinstein G.M., Weisenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of Drosophila melanogaster";  
PL Science 287:2185-2195(2000).  
CC -!- FUNCTION: HEPARAN SULFATE CELL SURFACE PROTEOGLYCAN. REQUIRED FOR  
CC CELL DIVISION PATTERNING DURING POST-EMBRYONIC DEVELOPMENT OF THE  
CC NERVOUS SYSTEM.  
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
CC (potential).  
CC -!- SIMILARITY: BELONGS TO THE GLYPICAN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; U31985; AAA97401.1; -;  
CC EMBL; AE003554; AAF50358.1; -;  
CC FlyBase; FB00011577; dally.  
CC InterPro; IPR001863; Glypican.  
CC Pfam; PF01533; Glypican; 1.  
CC PROSITE; PS01207; GLYPICAN; FALSE NEG.  
KW Proteoglycan; Heparan sulfate; Glycoprotein; Signal; GPI-anchor.  
FT SIGNAL 1 26 POTENTIAL.  
FT CHAIN 27 ? DIVISION ABNORMALLY DELAYED PROTEIN.  
FT PROPEP 626 626 REMOVED IN MATURE FORM (POTENTIAL).  
FT CARBOHYD 101 101 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 150 150 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 549 549 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).  
FT CARBOHYD 569 569 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).  
FT CARBOHYD 573 573 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).  
FT CARBOHYD 601 601 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).  
SQ SEQUENCE 626 AA; 69031 MW; 1182AD8A0DD0E4DD3 CRC64;  
  
Query Match 64.4%; Score 38; DB 1; Length 626;  
Best Local Similarity 41.7%; Pred. No. 3.4;  
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 CGXPAIXPXXTG 12  
Db 406 CQTESLTPYSSG 417  
|||:|:|:  
ID -CTRL\_HUMAN STANDARD; PRT; 264 AA.  
RESULT 10  
CTRL\_HUMAN  
AC P40313;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Chymotrypsin-like protease CTRL-1 precursor (EC 3.4.21.-).  
GN CTRL OR CTRL1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;

Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RA SEQUENCE FROM N.A.  
 RP MEDLINE=99000848; PubMed=9782094;  
 RX Pappano W.N., Scott I.C., Clark T.G., Eddy R.L., Shows T.B.,  
 RA Greenspan D.S.;  
 RT "Coding sequence and expression patterns of mouse chordin and mapping  
 of the cognate mouse chrd and human CHRD genes.";  
 RL Genomics 52:236-239(1998).  
 CC -1- FUNCTION: Dorsalizing factor. Key developmental protein that  
 dorsalizes early vertebrate embryonic tissues by binding to  
 and ventralizing TGF-beta family bone morphogenetic proteins (BMPs)  
 and sequestering them in latent complexes.  
 CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Detected at high levels of a in 7-day  
 postcoitum mouse embryos; its level decrease at later  
 developmental stages and in adult tissues.  
 CC -1- SIMILARITY: BELONGS TO THE CHORDIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 4 WFVC DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AF096276; AAD19895.1; -;  
 DR EMBL; AF069501; AAC68867.1; -;  
 DR MGD; MGI:1313268; Chrd  
 DR InterPro; IPR001007; VWF\_C.  
 DR Pfam; PF00093; vwc; 4.  
 DR SMART; SM00214; VWC; 4.  
 DR PROSITE; PS01208; VWF\_C; 2.  
 KW Developmental protein; Repeat; Glycoprotein; Signal.  
 FT SIGNAL 1 26 POTENTIAL.  
 FT CHAIN 27 948 CHORDIN.  
 FT DOMAIN 51 125 WFVC 1.  
 FT DOMAIN 701 758 WFVC 2.  
 FT DOMAIN 779 845 WFVC 3.  
 FT DOMAIN 867 927 WFVC 4.  
 FT CARBOHYD 347 347 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 877 877 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 948 AA; 101512 MW; 4DC2DA01D9BD2147 CRC64;  
 Query Match 59.3%; Score 35; DB 1; Length 948;  
 Best Local Similarity 50.0%; Pred. No. 21;  
 Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 CCXPALXPXTG 12  
 Db 407 CGADALIPVQTG 418  
 RESULT 12  
 CHRD\_HUMAN  
 ID CHRD\_HUMAN STANDARD; PRT; 955 AA.  
 AC Q9H2X0; Q9P022; Q9P023; Q9P024; Q9P025; Q9H2W8; Q9H2W9; Q9H2D3;  
 AC Q9S254;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Chordin precursor.  
 GN CHRD.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 4).  
 RX MEDLINE=21366001; PubMed=11472837;

Millet C., Lemaire P., Orsetti B., Guglielmi P., Francois V.;  
 "The human chordin gene encodes several differentially expressed  
 spliced variants with distinct BMP opposing activities.";  
 Mech. Dev. 106:85-96(2001).  
 [2]  
 RP SEQUENCE OF 115-955 FROM N.A. (ISOFORM 5).  
 RX MEDLINE=99000848; PubMed=9782094;  
 RA Pappano W.N., Scott I.C., Clark T.G., Eddy R.L., Shows T.B.,  
 RA Greenspan D.S.;  
 RT "Coding sequence and expression patterns of mouse chordin and mapping  
 of the cognate mouse chrd and human CHRD genes.";  
 RL Genomics 52:236-239(1998).  
 CC -1- FUNCTION: Dorsalizing factor. Key developmental protein that  
 dorsalizes early vertebrate embryonic tissues by binding to  
 and ventralizing TGF-beta family bone morphogenetic proteins (BMPs)  
 and sequestering them in latent complexes (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).  
 CC -1- ALTERNATIVE PRODUCTS: 5 isoforms; 1 (shown here), 2, 3, 4 and 5;  
 may be produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Expressed at the highest level in liver.  
 CC -1- SIMILARITY: BELONGS TO THE CHORDIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 4 WFVC DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AF209928; AAG35767.1; -;  
 DR EMBL; AF209929; AAG35768.1; -;  
 DR EMBL; AF209930; AAG35769.1; -;  
 DR EMBL; AF076612; AAC69835.1; -;  
 DR EMBL; AF283325; AAG35784.1; -;  
 DR EMBL; AF136632; AAF70236.1; -;  
 DR EMBL; AF136633; AAF70237.1; -;  
 DR EMBL; AF136634; AAF70238.1; -;  
 DR EMBL; AF136635; AAF70239.1; -;  
 DR Genew; HGNC:1949; CHRD.  
 DR MIM; 603475; -;  
 DR InterPro; IPR001007; VWF\_C.  
 DR Pfam; PF00093; vwc; 4.  
 DR SMART; SM00214; VWC; 4.  
 DR PROSITE; PS01208; VWF\_C; 2.  
 KW Developmental protein; Repeat; Glycoprotein; Signal;  
 KW Alternative splicing.  
 FT SIGNAL 1 26 POTENTIAL.  
 FT CHAIN 27 955 CHORDIN.  
 FT DOMAIN 51 125 WFVC 1.  
 FT DOMAIN 705 762 WFVC 2.  
 FT DOMAIN 784 850 WFVC 3.  
 FT DOMAIN 872 932 WFVC 4.  
 FT CARBOHYD 217 217 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 85 86 PQ -> GP (IN ISOFORM 2).  
 FT VARSPLIC 87 955 MISSING (IN ISOFORM 2).  
 FT VARSPLIC 85 94 PQMRRTRGP -> TGTLPREM (IN ISOFORM 3).  
 FT VARSPLIC 95 955 MISSING (IN ISOFORM 3).  
 FT VARSPLIC 328 350 GLTQVPLRLQLHQGLLELQA -> DSTPTGAARTSGQ  
 FT VARSPLIC 351 955 MISSING (IN ISOFORM 4).  
 FT VARSPLIC 441 480 MISSING (IN ISOFORM 4).  
 FT CONFLICT 115 118 RQLP -> QVAA (IN REF. 2).

[illegible]

FT CONFLICT 189 189 V -> A (IN REF. 2).  
FT CONFLICT 216 216 S -> P (IN REF. 2).  
FT CONFLICT 674 674 T -> P (IN REF. 2).  
SQ SEQUENCE 955 AA; 102013 MW; 12AC030CEACFF3ED CRC64;  
  
Query Match 59.3%; Score 35; DB 1; Length 955;  
Best Local Similarity 50.0%; Pred. No. 22;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGXPALPXVTG 12  
DB 411 CGADALIPVQTG 422

RESULT 13  
SERI DROME STANDARD; PRT; 265 AA.  
ID SERI DROME STANDARD; PRT; 265 AA.  
AC PI7205; Q9VAE0;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Serine proteases 1/2 precursor (EC 3.4.21.-)  
GN (SER99DA OR SER1 OR C9T877) AND (SER99DB OR SER2 OR C015519).  
OC Drosophila melanogaster (fruit fly)  
OS Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
ON NCBI\_TaxID=7227;  
RX RP SEQUENCE FROM N.A.  
RP MEDLINE=9219063; PubMed=2469005;  
RA Yun Y., Davis R.L.;  
RT "Levels of RNA from a family of putative serine protease genes are reduced in Drosophila melanogaster dunc mutants and are regulated by cyclic AMP.";  
RL Mol. Cell. Biol. 9:692-700(1989).  
RL [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.C., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H.C., Blazer B.G., Champe W., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.J., Heit G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burkitt K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo J.S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,  
RA Foster K.J., Gabrielian A.E., Garg N.S., Galbart W.M., Glaeser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphree L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith I.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Swirekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

YE Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
Zhang X.H., Zhong F.N., Zhou X., Zhu S., Zhu X., Smith H.O.,  
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
The genome sequence of Drosophila melanogaster";  
Science 287:2185-2195(2000).  
CC -!- FUNCTION: MAJOR FUNCTION MAY BE TO AID IN DIGESTION.  
CC -!- TISSUE SPECIFICITY: ABUNDANTLY EXPRESSED IN THE LARVAL GUT.  
CC -!- DEVELOPMENTAL STAGE: BEGAN TO APPEAR AT LATE EMBRYO STAGE AND CONTINUED TO INCREASE IN ABUNDANCE THROUGHOUT THE LARVAL STAGE. THEY ARE NOT PRESENT IN PUPAE BUT REAPPEARED IN THE ADULT.  
CC -!- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF SER1.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous gene model prediction.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to a license@isb-sib.ch).

EMBL; M24379; AAC02552.1;  
EMBL; M24379; AAC02553.1;  
EMBL; AE003771; AAF56971.1; ALT\_SEQ.  
EMBL; AE003771; AAF56972.1;  
PIR; J50260; J50260.  
PIR; A38894; A38894.  
HSP; P00761; LEPT.  
MEROPS; S01.UPA;  
FlyBase; FBgn0003356; Ser99Da.  
InterPro; IPR0003357; Ser99Db.  
InterPro; IPR001314; Chymotrypsin.  
InterPro; IPR00126; Ser proteas V8.  
InterPro; IPR001254; Ser-protease Try.  
Pfam; PF00089; trypsin\_1.  
PRINTS; PR00722; CHYMOTRYPSIN.  
PRINTS; PR00839; V8PROTEASE.  
SMART; SM00020; TRYPSIN\_SPC;  
PROSITE; PS02040; TRYPsin DOM; 1.  
PROSITE; PS00134; TRYPsin HIS; 1.  
PROSITE; PS00135; TRYPsin\_SER; 1.  
Hydrolase; Serine protease; Signal; Zymogen; Multigene family.  
FT SIGNAL 1 21 PROBABLE.  
FT PROPEP 22 35 BY SIMILARITY.  
FT CHAIN 36 265 SERINE PROTEASES 1/2.  
FT ACT SITE 78 78 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT SITE 123 123 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT SITE 215 215 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 63 79 BY SIMILARITY.  
FT DISULFID 189 201 BY SIMILARITY.  
FT DISULFID 211 239 BY SIMILARITY.  
FT CONFLICT 14 14 A -> T (IN REF. 1; AAB02553).  
SQ SEQUENCE 265 AA; 44848C523F0384B CRC64;

Query Match 57.6%; Score 34; DB 1; Length 265;  
Best Local Similarity 50.0%; Pred. No. 9.7;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 GXPALPXVTGY 13  
DB 242 GAPAVFSRVTYG 253

RESULT 14  
PPNK HALN1 STANDARD; PRT; 282 AA.  
ID PPNK HALN1 STANDARD; PRT; 282 AA.  
AC Q9HNX7;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Probable inorganic polyphosphate/ATP-NAD kinase (EC 2.7.1.23)

```

DE (Poly(P)/ATP NAD kinase).
GN PPNK OR VNG1900C.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Melti R., Goo Y.A.,
RA Leitbauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC -!- FUNCTION: Catalyzes the phosphorylation of NAD to NADP. Utilizes
CC ATP and other nucleoside triphosphates as well as inorganic
CC polyphosphate as a source of phosphorus (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + NAD(+) = ADP + NADP(+).
CC -!- COFACTOR: Requires divalent metal ions for activity (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE NAD KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X02138; CAA26064.2; -
CC EMBL; X00428; CAA25126.1; -
CC EMBL; L00036; AAA96678.1; -
CC PIR; A05242; Q0BE07.
CC InterPro; IPR000087; Collagen.
CC InterPro; IPR000714; BHV Unk.
CC Pfam; PF02053; Gene66; 1.
CC Zinc-finger.
CC ZN_FING 271 293
CC SQ SEQUENCE 312 AA; 34055 MW; 12D01B0E7C920EA3 CRC64;
CC
CC Query Match 57.6%; Score 34; DB 1; Length 282;
CC Best Local Similarity 50.0%; Pred. No. 10;
CC Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
CC
QY 1 CGXPAIXPXTG 12
DB 207 CGADALPPLVTG 218
|||:|
|||:|

RESULT 15
US10_HSV11 STANDARD; PRT; 312 AA.
ID US10_HSV11 STANDARD; PRT; 312 AA.
AC P06486;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Virion protein US10.
GN US10.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85160822; PubMed=2984429;
RA McGeoch D.J., Dolan A., Donald S., Rixon F.J.;
RT "Sequence determination and genetic content of the short unique
RT region in the genome of herpes simplex virus type 1.";
RL J. Mol. Biol. 181:1-13(1985).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=84169548; PubMed=6324121;
RA Rixon F.J., McGeoch D.J.;
RT "A 3' co-terminal family of mRNAs from the herpes simplex virus type
RT 1 short region: two overlapping reading frames encode unrelated
RT polypeptide one of which has highly reiterated amino acid sequence.";
RL Nucleic Acids Res. 12:2473-2487(1984).
CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUP TOGETHER HSV-1 US10,
CC BHV-1 66, EBV-4 ORF3, AND VZV 64/69.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X14112; CAA32275.1; -
CC EMBL; X02138; CAA26064.2; -
CC EMBL; X00428; CAA25126.1; -
CC EMBL; L00036; AAA96678.1; -
CC PIR; A05242; Q0BE07.
CC InterPro; IPR000087; Collagen.
CC InterPro; IPR000714; BHV Unk.
CC Pfam; PF02053; Gene66; 1.
CC Zinc-finger.
CC ZN_FING 271 293
CC SQ SEQUENCE 312 AA; 34055 MW; 12D01B0E7C920EA3 CRC64;
CC
CC Query Match 57.6%; Score 34; DB 1; Length 312;
CC Best Local Similarity 62.5%; Pred. No. 11;
CC Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
CC
QY 1 CGXPAIXP 8
DB 302 CGDPALTP 309
|||:|
|||:|

Search completed: February 12, 2003, 10:23:14
Job time : 3.74627 sec

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:16:06 ; Search time 9.19701 Seconds  
(without alignments)  
291.248 Million cell updates/sec

Title: US-10-036-371-3

Perfect score: 59

Sequence: 1 CGXPAIXPXTGY 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_nhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	88.1	23	13	Q9PR50 gadus morhu
2	52	88.1	24	13	Q9PR51 gadus morhu
3	52	88.1	263	13	Q9PWQ6 gadus morhu
4	51	86.4	260	13	Q9W7Q3 paralichthy
5	44	74.6	261	13	Q9W7Q4 paralichthy
6	44	74.6	263	11	Q9DC86 mus musculu
7	44	74.6	263	11	Q9DAX8 mus musculu
8	44	74.6	263	11	Q9CR35 mus musculu
9	43	72.9	1365	10	Q9ZP14 voivox cart
10	38	64.4	623	5	Q966V5 drosophila
11	37	62.7	103	11	Q9JJF2 mus musculu
12	37	62.7	542	16	Q9HTB5 pseudomonas
13	36	61.0	264	11	Q9D960 mus musculu
14	36	61.0	264	11	Q9D7P8 mus musculu
15	36	61.0	264	11	Q9EQ28 rattus norv
16	36	61.0	264	11	Q9ER05 mus musculu

17	36	61.0	321	10	Q41412
18	36	61.0	350	16	Q8UKL4
19	36	61.0	373	16	Q8YGV9
20	36	61.0	376	8	Q8XR04
21	36	61.0	376	8	Q8XR03
22	36	61.0	376	8	Q8XR02
23	36	61.0	376	10	Q65774
24	36	61.0	376	10	Q65777
25	36	61.0	485	12	Q8QUV3
26	36	61.0	840	12	Q8DGY8
27	36	61.0	840	12	Q8E9B4
28	36	61.0	842	12	Q69602
29	36	61.0	843	12	Q9DH92
30	36	61.0	843	12	Q9DH57
31	36	61.0	843	12	Q9E9B2
32	36	61.0	843	12	Q9E9A8
33	36	61.0	843	12	Q9E9A5
34	36	61.0	843	12	Q9E9A1
35	36	61.0	843	12	Q9E9A0
36	36	61.0	843	12	Q99HS9
37	36	61.0	843	12	Q99HR5
38	36	61.0	940	13	O57465
39	35	59.3	40	13	O9PSP2
40	35	59.3	111	17	O57786
41	35	59.3	120	17	O59100
42	35	59.3	189	1	Q52046
43	35	59.3	333	16	Q93RX9
44	35	59.3	351	3	Q9C0Y3
45	35	59.3	385	2	Q9X2J3

#### ALIGNMENTS

#### RESULT 1

Q9PR50  
ID Q9PR50 PRELIMINARY; PRT; 23 AA.  
AC Q9PR50;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE Chymotrypsin isoenzyme CHT2 (Fragment).  
OS Gadus morhua (Atlantic cod).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.  
OX NCBI\_TaxID=8049;  
RN [1]  
RP SEQUENCE  
RX MEDLINE=96043258; PubMed=7584866;  
RA Raai A.J., Flengsrud R., Sletten K.;  
RT "Chymotrypsin isoenzymes in Atlantic cod; differences in kinetics and substrate specificity."; RT  
RL Comp. Biochem. Physiol. 112B:393-398(1995).  
FT NON\_TER 1 1  
FT NON\_TER 23 23  
SQ SEQUENCE 23 AA; 2356 MW; ASEC2A7B73EB3F5E CRC64;

Query Match 88.1%; Score 52; DB 13; Length 23;  
Best Local Similarity 69.2%; Pred. No. 0.001;  
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGXPAIXPXTGY 13

Db 1 CGRPAISPVITGY 13

#### RESULT 2

Q9PR51  
ID Q9PR51 PRELIMINARY; PRT; 24 AA.  
AC Q9PR51;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

```

DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
DE Chymotrypsin isoenzyme CHT1 (Fragment).
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.
OX NCBI_TaxID=8049;
RN [1]
RP SEQUENCE.
RX MEDLINE=96043258; PubMed=7584866;
RA Rase A.J., Flensburg R., Sletten K.;
RT "Chymotrypsin isoenzymes in Atlantic cod; differences in kinetics and
RL substrate specificity.";
RL Comp. Biochem. Physiol. 112B:393-398 (1995).
DR HSSP; P00766; 1ACH.
FT NON_TER 1
FT TER 24
SQ SEQUENCE 24 AA; 2400 MW; 0A416ACA7B67F68D CRC64;

Query Match 88.1%; Score 52; DB 13; Length 24;
Best Local Similarity 69.2%; Pred. No. 0.0011;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CGXPAIXPXTGY 13
||| ||| |||
Db 1 CGNPAIQPVSTGY 13

RESULT 3
Q9PWQ6 PRELIMINARY; PRT; 263 AA.
AC Q9PWQ6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Chymotrypsin B precursor (EC 3.4.21.1).
GN CHYB.
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.
OX NCBI_TaxID=8049;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=PYLORIC CAECA;
RX MEDLINE=20464334; PubMed=11011764;
RA Spilliaert R., Gudmundsdottir A.;
RT "Molecular Cloning of the Atlantic Cod Chymotrypsinogen B.";
RL Microb. Comp. Genomics 5:41-50 (2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AJ242521; CAB43766.1; -.
DR HSSP; P00766; 1CHG.
DR MEROPS; S01.152; -.
DR InterPro; IPR001314; Chymotrypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Signal.
FT CHAIN 17 263 CHYMOTRYPSIN B.
FT SIGNAL 16 POTENTIAL.
FT TAG 17 263 CHYMOTRYPSIN B.
SQ SEQUENCE 263 AA; 28175 MW; EF61B18A34EE5E7C CRC64;

Query Match 88.1%; Score 52; DB 13; Length 263;
Best Local Similarity 69.2%; Pred. No. 0.012;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CGXPAIXPXTGY 13
||| ||| |||

```

```

Db 17 CGSPAIPQVVTGY 29

RESULT 4
Q9W7Q3 PRELIMINARY; PRT; 260 AA.
AC Q9W7Q3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Chymotrypsinogen 2.
OS Paralichthys olivaceus (Flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidei; Paralichthyidae; Paralichthys.
OX NCBI_TaxID=8255;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=PANCREAS;
RA Suzuki T., Srivastava A.S., Kurokawa T.;
RT "Japanese flounder mRNA for chymotrypsinogen 2.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB029754; BAA82366.1; -.
DR HSSP; P00766; 1CHG.
DR MEROPS; S01.152; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 260 AA; 27793 MW; 9F583044E22F78C0 CRC64;

Query Match 86.4%; Score 51; DB 13; Length 260;
Best Local Similarity 69.2%; Pred. No. 0.018;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CGXPAIXPXTGY 13
||| ||| |||
Db 16 CGSPAIPPVITGY 28

RESULT 5
Q9W7Q4 PRELIMINARY; PRT; 261 AA.
AC Q9W7Q4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Chymotrypsinogen 1.
OS Paralichthys olivaceus (Flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidei; Paralichthyidae; Paralichthys.
OX NCBI_TaxID=8255;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=PANCREAS;
RA Suzuki T., Srivastava A.S., Kurokawa T.;
RT "Japanese flounder mRNA for chymotrypsinogen 1.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB029753; BAA82365.1; -.
DR HSSP; P00763; 1DPO.
DR MEROPS; S01.256; -.

```

```

DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; Tryp_SPC; 1.
DR PROSITE: PS0240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 261 AA; 28184 MW; D7090A9D65395B7D CRC64;

Query Match 74.6%; Score 44; DB 13; Length 261;
Best Local Similarity 53.8%; Pred. No. 0.5;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 13
||| ||| |||
DB 17 CGVPSIKQVSSG 29

RESULT 6
Q9DC86 PRELIMINARY; PRT; 263 AA.
AC Q9DC86;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 2200008D09Rik protein.
GN 2200008D09Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=SPLEEN;
EX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsuki S.,
RA Hayashizaki Y.;
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL: AK003060; BAB2539.1; -.
DR HSP: P00766; 1GCT.
DR MEROPS: S01.152; -.
DR MGD: MGI:1913723; 2200008D09Rik.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; Tryp_SPC; 1.
DR PROSITE: PS0240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 263 AA; 27821 MW; 2620A27A5B04D CRC64;

```

```

Query Match 74.6%; Score 44; DB 11; Length 263;
Best Local Similarity 66.7%; Pred. No. 0.5;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 12
||| ||| |||
DB 19 CGVPAIQPVLTG 30

RESULT 7
Q9D8X8 PRELIMINARY; PRT; 263 AA.
AC Q9D8X8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 2200008D09Rik protein.
GN 2200008D09Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
EX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsuki S.,
RA Hayashizaki Y.;
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL: AK007566; BAB25112.1; -.
DR HSP: P00766; 1GCT.
DR MEROPS: S01.152; -.
DR MGD: MGI:1913723; 2200008D09Rik.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; Tryp_SPC; 1.
DR PROSITE: PS0240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 263 AA; 27898 MW; C0638FB8F905A92F CRC64;

Query Match 74.6%; Score 44; DB 11; Length 263;
Best Local Similarity 66.7%; Pred. No. 0.5;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 12
||| ||| |||
DB 19 CGVPAIQPVLTG 30

```

```

RESULT 8
Q9CR35 PRELIMINARY; PRT; 263 AA.
AC Q9CR35;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 2200008D09Rik protein.
GN 2200008D09Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1
SEQUENCE FROM N.A.
RC STRAIN=CS7BL/6J; TISSUE=STOMACH, SPLEEN, AND PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirni L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszew-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayaishizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AK008927; BAB25971.1; -
DR EMBL; AK003079; BAB22553.1; -
DR EMBL; AK007765; BAB25241.1; -
DR EMBL; AK007815; BAB25280.1; -
DR EMBL; AK008729; BAB25861.1; -
DR EMBL; AK008888; BAB25954.1; -
DR HSP; P00766; 1GCT.
DR MEROPS; S01.152; -
DR MGD; MGI:1913723; 2200008D09Rik.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser. protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Src; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 263 AA; 27822 MW; 28C4487AF1A26B27 CRC64;

Query Match 74.6%; Score 44; DB 11; Length 263;
Best Local Similarity 66.7%; Pred. No. 0.5;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 12
DB 19 CGVPAIQPLVIG 30

RESULT 9
Q9ZP14 PRELIMINARY; PRT; 1365 AA.
ID Q9ZP14
AC Q9ZP14;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)

Query Match 64.4%; Score 38; DB 5; Length 623;
Best Local Similarity 41.7%; Pred. No. 20;
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 12
DB 403 CGTFLTPYSSG 414

RESULT 11
Q9JUF2 PRELIMINARY; PRT; 103 AA.
ID Q9JUF2
AC Q9JUF2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)

Query Match 64.4%; Score 38; DB 5; Length 623;
Best Local Similarity 41.7%; Pred. No. 20;
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 12
DB 403 CGTFLTPYSSG 414

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Alkaline phosphatase precursor (EC 3.1.3.1).
GN PHOX.
OS Volvox carteri.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Volvocaceae; Volvox.
OX NCBI_TaxID=3067;
RN 1
SEQUENCE FROM N.A.
RC STRAIN=HK 10;
RX MEDLINE=99098915; PubMed=9880549;
RA Hallmann A.;
RT "Enzymes in the extracellular matrix of Volvox: an inducible, calcium-
RT dependent phosphatase with a modular composition.";
RL J. Biol. Chem. 274:1691-1697(1999).
RW EMBL; AJ012458; CAA10030.1; -
KW Hydrolase; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 1365 ALKALINE PHOSPHATASE.
SQ SEQUENCE 1365 AA; 146283 MW; 0290DA5C4A792CC4 CRC64;

Query Match 72.9%; Score 43; DB 10; Length 1365;
Best Local Similarity 53.8%; Pred. No. 4.1;
Matches 7; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 13
DB 1063 CGTPTNPAAPGY 1075

RESULT 10
Q966V5 PRELIMINARY; PRT; 623 AA.
AC Q966V5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Dally.
GN DALLY OR CG4974.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN 1
SEQUENCE FROM N.A.
RC TEUDA M., Izumi S., Nakato H.;
RT "Transcriptional and Rosttranscriptional Regulation of the dally Gene,
RT a Drosophila Integral Membrane Proteoglycan.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB052367; BAB60703.1; -
DR EMBL; AB052365; BAB60703.1; JOINED.
DR EMBL; AB052366; BAB60703.1; JOINED.
DR FlyBase; FBgn0011577; dally.
DR InterPro; IPR001863; Glypican.
DR Pfam; PF01153; Glypican; 1.
SQ SEQUENCE 623 AA; 68791 MW; 89A8410D6BCE3411 CRC64;

Query Match 64.4%; Score 38; DB 5; Length 623;
Best Local Similarity 41.7%; Pred. No. 20;
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 12
DB 403 CGTFLTPYSSG 414

RESULT 11
Q9JUF2 PRELIMINARY; PRT; 103 AA.
ID Q9JUF2
AC Q9JUF2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)

```

```

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Brain cDNA, clone MNCB-0953, similar to Mus musculus SPAG4 protein
DE alternative spliced.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Osada N., Kusuda J., Tanuma R., Ito A., Hirata M., Sugano S.,
RA Hashimoto K.;
RT "Isolation of full-length cDNA clones from mouse brain cDNA library
RT made by oligo-capping method.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB041554; BAA95039.1; -
SQ SEQUENCE 103 AA; 10590 MW; E8707BEA59ED8372 CRC64;

Query Match 62.7%; Score 37; DB 11; Length 103;
Best Local Similarity 50.0%; Pred. No. 5.3;
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 12
DB 59 CGEPALSGMPG 70

RESULT 12
Q9HTB5 ID Q9HTB5 PRELIMINARY; PRT; 542 AA.
AC Q9HTB5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Hypothetical protein PA5455.
GN PA5455.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=2043737; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saler M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004958; AAG08840.1; -
DR InterPro: IPR001296; Glycos transf. 1.
DR Pfam: PF00534; Glycos transf. 1.
KW Hypothetical protein-Complete proteome.
SQ SEQUENCE 542 AA; 61018 MW; 537F8A978804CAB CRC64;

Query Match 62.7%; Score 37; DB 16; Length 542;
Best Local Similarity 75.0%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGXPAIXP 8
DB 420 CGTFAIAP 427

RESULT 13
Q9D960 ID Q9D960 PRELIMINARY; PRT; 264 AA.
AC Q9D960;

```

---

```

DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 1810004D15RIK protein.
GN CTRL OR 1810004D15RIK.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata Y., Yoshino M., Itoh M., Ishii Y.,
RA Akawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guncinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY. BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AK007333; BAB24967.1; -
DR HSPP; P00766; 4CHA.
DR MEROPS; S01.256; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser. protease_Try.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 264 AA; 28151 MW; 1D979719E07C16DE CRC64;

Query Match 61.0%; Score 16; DB 11; Length 264;
Best Local Similarity 75.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGXPAIXP 8
DB 19 CGVPAITP 26

RESULT 14
Q9D7P8 ID Q9D7P8 PRELIMINARY; PRT; 264 AA.
AC Q9D7P8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 1810004D15RIK protein.
GN CTRL OR 1810004D15RIK.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

```

RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=TONGUE;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Mateu Y., Nikaudo I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaudo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 TRYPsin FAMILY.  
 DR EMBL; AK009019; BAB26029.1; -.  
 DR HSSP; P00766; 4CHA.  
 DR MEROPS; S01.256; -.  
 DR MGD; MGI:88558; Ctr1.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPsin.  
 DR SMART; SM00020; Tryp\_SPC; 1.  
 DR PROSITE; PS02040; TRYPsin\_DOM; 1.  
 DR PROSITE; PS00134; TRYPsin\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00135; TRYPsin\_SER; 1.  
 KW Hydrolase; Serine protease.  
 SQ SEQUENCE 264 AA; 28116 MW; 1D979469A07056C2 CRC64;

Query Match 61.0%; Score 36; DB 11; Length 264;  
 Best Local Similarity 75.0%; Pred. No. 22;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGXPAIXP 8  
 |||||  
 Db 19 CGVPAITP 26

RESULT 15  
 Q9EQZ8 PRELIMINARY; PRT; 264 AA.  
 AC Q9EQZ8;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Chymopasin.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=RAT PANCREAS;  
 RA Sogame Y., Mitsui S., Kataoka K., Kashima K., Kato M., Sakagami J.,  
 RA Yamaguchi N.;  
 RT "Molecular cloning of rat chymopasin.";  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 TRYPsin FAMILY  
 CC EMBL; AB020757; BAB20287.1; -.  
 DR HSSP; P00766; 4CHA.  
 DR MEROPS; S01.256; -.  
 . . .

DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPsin.  
 DR SMART; SM00020; Tryp\_SPC; 1.  
 DR PROSITE; PS02040; TRYPsin\_DOM; 1.  
 DR PROSITE; PS00134; TRYPsin\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00135; TRYPsin\_SER; 1.  
 KW Hydrolase; Serine protease.  
 SQ SEQUENCE 264 AA; 28116 MW; F9ED5D210FD3500E CRC64;  
 Query Match 61.0%; Score 36; DB 11; Length 264;  
 Best Local Similarity 75.0%; Pred. No. 22;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 CGXPAIXP 8  
 |||||  
 Db 19 CGVPAITP 26  
 Search completed: February 12, 2003, 10:27:22  
 Job time : 13.197 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:18:40 ; Search time 2.79403 Seconds  
(without alignment)  
136.898 Million cell updates/sec

Title: US-10-036-371-3  
Perfect score: 59  
Sequence: 1 CGXPAIXPXXTG 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/prodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/prodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/prodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/prodata/1/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	35	59.3	855	2 US-08-938-365-3
2	35	59.3	867	2 US-08-938-365-2
3	35	59.3	954	2 US-08-749-169A-3
4	35	59.3	954	2 US-09-130-032A-3
5	34	57.6	662	4 US-09-134-001C-4074
6	33	55.9	15	4 US-09-009-953-163
7	33	55.9	845	1 US-08-416-950-11
8	33	55.9	845	2 US-08-469-830-11
9	33	55.9	845	6 5196194-17
10	32	54.2	106	1 US-08-491-976-1
11	32	54.2	106	3 US-08-476-705A-7
12	32	54.2	106	4 US-08-631-469B-20
13	32	54.2	106	4 US-09-056-868B-15
14	32	54.2	154	2 US-08-330-394B-29
15	32	54.2	156	2 US-08-330-394A-22
16	32	54.2	159	1 US-08-491-976-3
17	32	54.2	162	1 US-08-048-164A-2
18	32	54.2	162	1 US-08-460-462-2
19	32	54.2	162	1 US-08-460-457-2
20	32	54.2	162	1 US-08-460-458-2
21	32	54.2	162	2 US-08-460-455-2
22	32	54.2	162	2 US-08-330-394B-2
23	32	54.2	163	3 US-09-006-636-7
24	32	54.2	163	4 US-09-006-632-7
25	32	54.2	163	4 US-09-325-274-7
26	32	54.2	382	4 US-09-277-716-22
27	32	54.2	382	4 US-09-609-161B-22

28	32	54.2	2972	4	US-08-469-260A-387	Sequence 387, App
29	31	52.5	36	6	5202417-3	Patent No. 5202417
30	31	52.5	60	3	US-08-476-705A-5	Sequence 5, Appl
31	31	52.5	60	4	US-08-631-469B-15	Sequence 15, Appl
32	31	52.5	60	4	US-08-631-469B-19	Sequence 19, Appl
33	31	52.5	60	4	US-09-056-868B-10	Sequence 10, Appl
34	31	52.5	60	6	5202417-1	Patent No. 5202417
35	31	52.5	66	3	US-08-782-480-26	Sequence 26, Appl
36	31	52.5	66	4	US-08-954-211-26	Sequence 26, Appl
37	31	52.5	66	4	US-09-005-167A-26	Sequence 26, Appl
38	31	52.5	66	4	US-09-176-741B-26	Sequence 26, Appl
39	31	52.5	83	6	5514558-5	Patent No. 5514558
40	31	52.5	84	4	US-08-134-198E-5	Sequence 5, Appl
41	31	52.5	84	6	5202417-2	Patent No. 5202417
42	31	52.5	125	4	US-09-199-637A-317	Sequence 317, App
43	31	52.5	477	4	US-09-134-001C-3487	Sequence 3487, Ap
44	31	52.5	492	4	US-09-134-001C-4847	Sequence 4847, Ap
45	31	52.5	607	4	US-09-204-208A-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1  
US-08-938-365-3  
; Sequence 3, Application US/08938365  
; Patent No. 5989909  
; GENERAL INFORMATION:  
; APPLICANT: Yang, Pan  
; TITLE OF INVENTION: HUCHORDIN AND USES THEREOF  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/938,365  
; FILING DATE: 26-SEP-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meiklejohn Ph.D.; Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 09404/040001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 855 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-938-365-3

Query Match 59.3%; Score 35; DB 2; Length 855;  
Best Local Similarity 50.0%; Pred. No. 96;  
Matches 6; Conservative 1; Mismatches 5; Indels 3; Gaps 0;

Qy 1 CGXPAIXPXXTG 12

Db 406 CGADALIPVQTG 417

RESULT 2  
US-08-938-365-2  
; Sequence 2, Application US/08938365  
; Patent No. 5989909  
; GENERAL INFORMATION:  
; APPLICANT: Yang, Pan  
; TITLE OF INVENTION: HUCHORDIN AND USES THEREOF  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/938,365  
; FILING DATE: 26-SEP-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meiklejohn, Ph.D., Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 09404/040001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 867 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
; US-08-938-365-2  
  
Query Match 59.3%; Score 35; DB 2; Length 867;  
Best Local Similarity 50.0%; Pred. No. 98;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
  
Qy 1 CGXPAIXPXXTG 12  
|||:|:|  
Db 411 CGADALIPVQTG 422  
  
RESULT 3  
US-08-749-169A-3  
; Sequence 3, Application US/08749169A  
; Patent No. 5846770  
; GENERAL INFORMATION:  
; APPLICANT: RACIE, Lisa  
; APPLICANT: LAVALLIE, Edward  
; APPLICANT: DEROBERTIS, Edward  
; TITLE OF INVENTION: CHORDIN COMPOSITIONS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/130,032A  
; FILING DATE: August 4, 1998  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: LAZAR, Steven R.  
; REGISTRATION NUMBER: 32,618  
; REFERENCE/DOCKET NUMBER: GI 5284-DIV  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 498-8260  
; TELEFAX: (617) 876-5851  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 954 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-130-032A-3

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/749,169A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: LAZAR, Steven R.  
; REGISTRATION NUMBER: 32,618  
; REFERENCE/DOCKET NUMBER: GI 5284  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 498-8260  
; TELEFAX: (617) 876-5851  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 954 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-749-169A-3  
  
Query Match 59.3%; Score 35; DB 2; Length 954;  
Best Local Similarity 50.0%; Pred. No. 11e+02;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
  
Qy 1 CGXPAIXPXXTG 12  
|||:|:|  
Db 411 CGADALIPVQTG 422  
  
RESULT 4  
US-09-130-032A-3  
; Sequence 3, Application US/09130032A  
; Patent No. 5986056  
; GENERAL INFORMATION:  
; APPLICANT: LaVallie, Edward  
; APPLICANT: Racie, Lisa  
; APPLICANT: Derobertis, Edward  
; TITLE OF INVENTION: HUMAN CHORDIN COMPOSITIONS  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/130,032A  
; FILING DATE: August 4, 1998  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: LAZAR, Steven R.  
; REGISTRATION NUMBER: 32,618  
; REFERENCE/DOCKET NUMBER: GI 5284-DIV  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 498-8260  
; TELEFAX: (617) 876-5851  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 954 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-130-032A-3



```

Query Match          59.3%; Score 35; DB 2; Length 954;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CGXPAIXXTG 12
Db 411 CGADALIPVQTG 422

RESULT 5
US-09-134-001C-4074
; Sequence 4074, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4074
; LENGTH: 662
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4074

Query Match          57.6%; Score 34; DB 4; Length 662;
Best Local Similarity 60.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 4 PAIXPXTG 13
Db 294 PAIGTUSGY 303

RESULT 6
US-09-009-953-163
; Sequence 163, Application US/09009953
; Patent No. 6413517
; GENERAL INFORMATION:
; APPLICANT: Sette, Alessandro
; TITLE OF INVENTION: Identification of Broadly
; NUMBER OF SEQUENCES: 274
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009,953
; FILING DATE: 21-Jan-1998
; CLASSIFICATION: <Unknown>
; APPLICATION DATA:
; APPLICATION NUMBER: US 60/036,713
; FILING DATE: 23-JAN-1997
; APPLICATION NUMBER: US 60/037,432
; FILING DATE: 07-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762

```

```

; REFERENCE/DOCKET NUMBER: 018623-011520US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 163:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 163:
US-09-009-953-163

Query Match          55.9%; Score 33; DB 4; Length 15;
Best Local Similarity 62.5%; Pred. No. 4.7;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGXPAIXP 8
Db 2 CGYPALMP 9

RESULT 7
US-08-416-950-11
; Sequence 11, Application US/08416950
; Patent No. 5780036
; GENERAL INFORMATION:
; APPLICANT: CHISARI, Francis V.
; TITLE OF INVENTION: PEPTIDES FOR INDUCING CYTOTOXIC T
; TITLE OF INVENTION: LYMPHOCYTE RESPONSES TO HEPATITIS B VIRUS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Khourie and Crew
; STREET: One Market Plaza, Steuart Street Tower
; CITY: San Francisco
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/416,950
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US
; FILING DATE:
; APPLICATION NUMBER: US 07/935,898
; FILING DATE: 26-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/749,540
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 14740-2-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 467-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 845 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-416-950-11

```

Query Match 55.9%; Score 33; DB 1; Length 845;  
Best Local Similarity 62.5%; Pred. No. 2.2e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGXPAIXP 8  
DB 651 CGYPALMP 658

RESULT 8  
US-08-469-830-11  
; Sequence 11, Application US/08469830  
; Patent No. 5932224  
; GENERAL INFORMATION:  
; APPLICANT: CHIGARI, Francis V.  
; TITLE OF INVENTION: PEPTIDES FOR INDUCING CYTOTOXIC T  
; TITLE OF INVENTION: LYMPHOCYTE RESPONSES TO HEPATITIS B VIRUS  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Kourie and Crew  
; STREET: One Market Plaza, Steuart Street Tower  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 94105-1492  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/469,830  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/100,870  
; FILING DATE: 02-AUG-1993  
; APPLICATION NUMBER: US 07/935,898  
; FILING DATE: 26-AUG-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/749,540  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parmelee, Steven W.  
; REGISTRATION NUMBER: 31,990  
; REFERENCE/DOCKET NUMBER: 14740-2-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 467-9600  
; TELEFAX: (415) 543-5043  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 845 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-469-830-11

Query Match 55.9%; Score 33; DB 2; Length 845;  
Best Local Similarity 62.5%; Pred. No. 2.2e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGXPAIXP 8  
DB 651 CGYPALMP 658

RESULT 9  
5196194-17  
; Patent No. 5196194  
; APPLICANT: RUTTER, WILLIAM J.; GOODMAN, HOWARD M.  
; TITLE OF INVENTION: VACCINES CONTAINING HEPATITIS B S-PROTEIN  
; NUMBER OF SEQUENCES: 21

Query Match 55.9%; Score 33; DB 6; Length 845;  
Best Local Similarity 62.5%; Pred. No. 2.2e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGXPAIXP 8  
DB 651 CGYPALMP 658

RESULT 10  
US-08-491-976-1  
; Sequence 1, Application US/08491976  
; Patent No. 5783416  
; GENERAL INFORMATION:  
; APPLICANT: Thim, Lars  
; APPLICANT: No. 5783416ris, Kjeld  
; APPLICANT: No. 5783416ris, Fanny  
; APPLICANT: Bjorn, Soren E.  
; APPLICANT: Christensen, Mogens  
; APPLICANT: Nielsen, Per P.  
; TITLE OF INVENTION: Human Spasmodytic Polypeptide in  
; TITLE OF INVENTION: Glycosylated Form  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5783416o No. 5783416disk of No. 5783416th America, Inc.  
; STREET: 405 Lexington Avenue, 64th Floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 10174-6401  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/491,976  
; FILING DATE: 02-AUG-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lambiris, Elias J.  
; REGISTRATION NUMBER: 33,728  
; REFERENCE/DOCKET NUMBER: 3951.204-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 106 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-491-976-1

Query Match 54.2%; Score 32; DB 1; Length 106;  
Best Local Similarity 62.5%; Pred. No. 46;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGXPAIXP 8  
|||  
Db 68 CGYFGISP 75

RESULT 11

US-08-476-705A-7  
; Sequence 7, Application US/08476705A  
; Patent No. 6063755  
; GENERAL INFORMATION:  
; APPLICANT: Podolsky, Daniel K.  
; TITLE OF INVENTION: INTESTINAL TREFOIL PROTEINS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/476,705A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meiklejohn, Ph.D., Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 00786/066004  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-542-5070  
; TELEFAX: 617-542-8906  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 106 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-476-705A-7

Query Match 54.2%; Score 32; DB 3; Length 106;  
Best Local Similarity 62.5%; Pred. No. 46;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGXPAIXP 8  
|||  
Db 68 CGYFGISP 75

RESULT 12

US-08-631-469B-20  
; Sequence 20, Application US/08631469B  
; Patent No. 6221840  
; GENERAL INFORMATION:  
; APPLICANT: Daniel K. Podolsky  
; TITLE OF INVENTION: INTESTINAL TREFOIL PROTEINS  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/631,469B  
FILING DATE: 12-APR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/631,469  
FILING DATE: 12-APR-1996  
APPLICATION NUMBER: 08/191,352  
FILING DATE: 02-FEB-1994  
APPLICATION NUMBER: 08/037,741  
FILING DATE: 25-MAR-1993  
APPLICATION NUMBER: 07/837,192  
FILING DATE: 13-FEB-1992  
APPLICATION NUMBER: 07/655,965  
FILING DATE: 14-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Meiklejohn, Ph.D., Anita L.  
REGISTRATION NUMBER: 35,283  
REFERENCE/DOCKET NUMBER: 00786/322001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
TELEX: 200107  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 20:  
US-08-631-469B-20

Query Match 54.2%; Score 32; DB 4; Length 106;  
Best Local Similarity 62.5%; Pred. No. 46;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGXPAIXP 8  
|||  
Db 68 CGYFGISP 75

RESULT 13

US-09-056-868B-15  
; Sequence 15, Application US/09056868B  
; Patent No. 6316218  
; GENERAL INFORMATION:  
; APPLICANT: Podolsky, Daniel K.  
; TITLE OF INVENTION: INTESTINAL TREFOIL PROTEINS  
; FILE REFERENCE: 00786-066005  
; CURRENT APPLICATION NUMBER: US/09/056,868B  
; CURRENT FILING DATE: 1998-01-27  
; PRIOR APPLICATION NUMBER: US 08/476,705  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: US 08/191,352  
; PRIOR FILING DATE: 1994-02-02  
; PRIOR APPLICATION NUMBER: US 08/037,741  
; PRIOR FILING DATE: 1993-03-25  
; PRIOR APPLICATION NUMBER: US 07/837,192  
; PRIOR FILING DATE: 1992-02-13  
; PRIOR APPLICATION NUMBER: US 07/655,965  
; PRIOR FILING DATE: 1991-02-14  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15  
; LENGTH: 106  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetically generated peptide  
US-09-056-868B-15

Query Match 54.2%; Score 32; DB 4; Length 106;  
Best Local Similarity 62.5%; Pred. No. 46;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 CGXPAIXP 8  
68 CGYFGISP 75

RESULT 14  
US-08-330-394A-29  
; Sequence 29, Application US/08330394A  
; Patent No. 5856201  
; GENERAL INFORMATION:  
; APPLICANT: Shoseyov, Oded  
; APPLICANT: Yosef, Karmey  
; APPLICANT: Shpiegl, Itai  
; APPLICANT: Goldstein, Marc A.  
; APPLICANT: Dol, Roy H.  
; TITLE OF INVENTION: METHODS OF DETECTION USING THE  
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 AVENUE OF THE AMERICAS  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/330.394A  
; FILING DATE: 27-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: CIP OF PCT/US94/04132  
; FILING DATE: 14-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MISROCK, S. LESLIE  
; REFERENCE/DOCKET NUMBER: 7809-005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864  
; TELEX: 66441 PENNIE  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 156 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-330-394A-22

Query Match 54.2%; Score 32; DB 2; Length 156;  
Best Local Similarity 50.0%; Pred. No. 67;  
Matches 5; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 4 PAIXPXXTGY 13  
| : | | |  
Db 136 PVVNPXVTGY 145

Search completed: February 12, 2003, 10:30:08  
Job time : 4.79403 secs

1 CGXPAIXP 8  
68 CGYFGISP 75

RESULT 15  
US-08-330-394A-22  
; Sequence 22, Application US/08330394A  
; Patent No. 5856201  
; GENERAL INFORMATION:  
; APPLICANT: Shoseyov, Oded  
; APPLICANT: Yosef, Karmey  
; APPLICANT: Shpiegl, Itai  
; APPLICANT: Goldstein, Marc A.  
; APPLICANT: Dol, Roy H.  
; TITLE OF INVENTION: METHODS OF DETECTION USING THE  
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 AVENUE OF THE AMERICAS  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/330.394A  
; FILING DATE: 27-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: CIP OF PCT/US94/04132  
; FILING DATE: 14-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MISROCK, S. LESLIE  
; REFERENCE/DOCKET NUMBER: 7809-005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864  
; TELEX: 66441 PENNIE  
; INFORMATION FOR SEQ ID NO: 29:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 154 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-330-394A-29

Query Match 54.2%; Score 32; DB 2; Length 154;  
Best Local Similarity 50.0%; Pred. No. 66;  
Matches 5; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 4 PAIXPXXTGY 13  
| : | | |  
Db 136 PVVNPXVTGY 145

RESULT 15  
US-08-330-394A-22  
; Sequence 22, Application US/08330394A  
; Patent No. 5856201  
; GENERAL INFORMATION:  
; APPLICANT: Shoseyov, Oded  
; APPLICANT: Yosef, Karmey  
; APPLICANT: Shpiegl, Itai  
; APPLICANT: Goldstein, Marc A.  
; APPLICANT: Dol, Roy H.  
; TITLE OF INVENTION: METHODS OF DETECTION USING THE  
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN  
; NUMBER OF SEQUENCES: 33

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:22:36 ; Search time 1.9791 Seconds

(without alignments)  
167.821 Million cell updates/sec

Title: US-10-036-371-3

Perfect score: 59

Sequence: 1 CXSPAIXPXTGY 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US03\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51	86.4	13	12	US-10-036-371-3
2	40	67.8	13	12	US-10-036-371-3
3	40	67.8	146	10	US-09-925-297-643
4	40	67.8	263	10	US-09-888-615-96
5	35	59.3	192	10	US-09-925-297-529
6	35	59.3	867	9	US-09-811-088-10
7	35	59.3	948	10	US-09-897-699-2
8	35	59.3	954	9	US-09-944-413-7
9	35	59.3	954	9	US-09-944-403-7
10	35	59.3	954	9	US-09-944-896-7
11	35	59.3	954	9	US-09-944-944-7
12	35	59.3	954	9	US-09-944-907-7
13	35	59.3	954	9	US-09-944-929-7
14	35	59.3	954	10	US-09-866-028-7
15	35	59.3	954	10	US-09-944-449-7
16	35	59.3	954	10	US-09-944-457-7
17	35	59.3	954	10	US-09-944-862-7
18	35	59.3	954	10	US-09-945-587-7
19	35	59.3	954	10	US-09-945-015-7

20	35	59.3	954	10	US-09-944-396-7
21	35	59.3	954	10	US-09-944-097-7
22	35	59.3	954	10	US-09-944-432-7
23	35	59.3	954	10	US-09-943-762-7
24	35	59.3	954	10	US-09-944-654-7
25	35	59.3	954	10	US-09-943-851A-7
26	35	59.3	955	9	US-10-044-716-8
27	35	59.3	955	9	US-09-798-051-8
28	35	57.6	200	10	US-09-925-299-817
29	34	57.6	360	9	US-09-905-291A-213
30	34	57.6	360	9	US-09-902-853-213
31	34	57.6	360	9	US-09-907-824-213
32	34	57.6	360	9	US-09-907-841-213
33	34	57.6	360	9	US-09-904-011-213
34	34	57.6	360	9	US-10-174-590-28
35	34	57.6	360	9	US-10-176-758-28
36	34	57.6	360	9	US-10-175-737-28
37	34	57.6	360	9	US-09-906-742-213
38	34	57.6	360	9	US-10-173-706-28
39	34	57.6	360	9	US-10-175-738-28
40	34	57.6	360	9	US-10-175-752-28
41	34	57.6	360	9	US-10-176-482-28
42	34	57.6	360	9	US-10-176-757-28
43	34	57.6	360	9	US-10-176-913-28
44	34	57.6	360	9	US-10-180-552-28
45	34	57.6	360	9	US-10-180-557-28

## ALIGNMENTS

RESULT 1  
US-10-036-371-3  
; Sequence 3, Application US/10036371  
; Patent No. US20020141987A1  
; GENERAL INFORMATION:  
; APPLICANT: BJARNARSON, JON B.  
; TITLE OF INVENTION: FISH SERINE PROTEINASES AND THEIR PHARMACEUTICAL AND  
; FILE OF INVENTION: COSMETIC USE  
; FILE REFERENCE: 81691/284960  
; CURRENT APPLICATION NUMBER: US/10/036,371  
; PRIOR FILING DATE: 2002-01-07  
; PRIOR APPLICATION NUMBER: 09/411,688  
; PRIOR FILING DATE: 1999-10-12  
; PRIOR APPLICATION NUMBER: 5086/99  
; PRIOR FILING DATE: 1999-06-18  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 13  
; TYPE: PRT  
; ORGANISM: Gadus sp.  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (3)  
; OTHER INFORMATION: R or S  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (7)  
; OTHER INFORMATION: S or Q  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (9)  
; OTHER INFORMATION: V or Q  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (10)  
; OTHER INFORMATION: I or V  
US-10-036-371-3

Query Match 86.4%; Score 51; DB 12; Length 13;  
Best Local Similarity 100.0%; Pred.No. 0.0005;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
;
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (137)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (143)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-643

Query Match          67.8%; Score 40; DB 10; Length 146;
Best Local Similarity 58.3%; Pred. No. 0.72;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CGXPAIXPXXTG 12
Db 23 CGVPAIHPVLGS 34
      |||||
      |||||

RESULT 4
US-09-888-615-96
; Sequence 96, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: WHYTE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARYDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 96
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-888-615-96

Query Match          67.8%; Score 40; DB 10; Length 263;
Best Local Similarity 58.3%; Pred. No. 1.3;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CGXPAIXPXXTG 12
Db 19 CGVPAIHPVLGS 30
      |||||
      |||||

RESULT 5
US-09-925-297-529
; Sequence 529, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: ROSEN ET AL.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 529
; LENGTH: 192
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-297-529
```

```
;
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (94)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (126)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (130)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (133)

US-10-036-371-5
; Sequence 5, Application US/10036371
; Patent No. US20020141987A1
; GENERAL INFORMATION:
; APPLICANT: BJARNARSON, JON B.
; TITLE OF INVENTION: FISH SERINE PROTEINASES AND THEIR PHARMACEUTICAL AND
; TITLE OF INVENTION: COSMETIC USE
; FILE REFERENCE: 81691/284960
; CURRENT APPLICATION NUMBER: US/10/036,371
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: 09/411,688
; PRIOR FILING DATE: 1999-10-12
; PRIOR APPLICATION NUMBER: 5086/99
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Bovine sp.
US-10-036-371-5

Query Match          67.8%; Score 40; DB 12; Length 13;
Best Local Similarity 58.3%; Pred. No. 0.067;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CGXPAIXPXXTG 12
Db 1 CGVPAIQPVLGS 12
      |||||
      |||||

RESULT 3
US-09-925-297-643
; Sequence 643, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: ROSEN ET AL.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 643
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (94)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (126)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (130)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (133)
```

```
Query Match      59.3%; Score 35; DB 10; Length 192;
Best Local Similarity 75.0%; Pred. No. 8, 7;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGXPAIXP 8
Db 15 CGIPAIXP 22

RESULT 6
US-09-811-088-10
; Sequence 10, Application US/09811088
; Patent No. US2002016046A1
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Gearing, David P.
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC AND OTHER
; FILE REFERENCE: 07334-324001
; CURRENT APPLICATION NUMBER: US/09/811,088
; PRIOR FILING DATE: 2001-03-16
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: US 08/820,364
; PRIOR FILING DATE: 1997-03-12
; PRIOR APPLICATION NUMBER: US 09/757,421
; PRIOR FILING DATE: 2001-01-10
; PRIOR APPLICATION NUMBER: US 08/843,652
; PRIOR FILING DATE: 1997-04-16
; PRIOR APPLICATION NUMBER: US 08/843,651
; PRIOR FILING DATE: 1997-04-16
; PRIOR APPLICATION NUMBER: US 09/354,809
; PRIOR FILING DATE: 1999-07-16
; PRIOR APPLICATION NUMBER: US 08/938,365
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 867
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-811-088-10

Query Match      59.3%; Score 35; DB 9; Length 867;
Best Local Similarity 50.0%; Pred. No. 38;
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 12
Db 411 CGADALIPVQTG 422

RESULT 7
US-09-897-699-2
; Sequence 2, Application US/09897699
; Patent No. US2002038468A1
; GENERAL INFORMATION:
; APPLICANT: Leventen, Michael W.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING CHORDIN GENE
; FILE REFERENCE: R-269
; CURRENT APPLICATION NUMBER: US/09/897,699
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215,179
; PRIOR FILING DATE: 2000-06-29
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 948
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-897-699-2

Query Match      59.3%; Score 35; DB 10; Length 948;
Best Local Similarity 50.0%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 12
Db 407 CGADALIPVQTG 418

RESULT 8
US-09-944-413-7
; Sequence 7, Application US/09944413
; Patent No. US2002015600A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Bolstein, David
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul
; APPLICANT: Grimaldi, Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Hillan, Kenneth
; APPLICANT: Kijavini, Ivar
; APPLICANT: Napier, Mary
; APPLICANT: Roy, Margaret
; APPLICANT: Tamas, Daniel
; APPLICANT: Wood, William
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P2548P1C1
; CURRENT APPLICATION NUMBER: US/09/944,413
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 09/866,028
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 60/067,411
; PRIOR FILING DATE: December 3, 1997
; PRIOR APPLICATION NUMBER: 60/069,334
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,335
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,278
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,425
; PRIOR FILING DATE: December 12, 1997
; PRIOR APPLICATION NUMBER: 60/069,696
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,694
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,702
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,870
; PRIOR FILING DATE: December 17, 1997
; PRIOR APPLICATION NUMBER: 60/069,873
; PRIOR FILING DATE: December 17, 1997
; PRIOR APPLICATION NUMBER: 60/068,017
; PRIOR FILING DATE: December 18, 1997
; PRIOR APPLICATION NUMBER: 60/070,440
; PRIOR FILING DATE: January 5, 1998
; PRIOR APPLICATION NUMBER: 60/074,086
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/074,092
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/075,945
; PRIOR FILING DATE: February 25, 1998
; PRIOR APPLICATION NUMBER: 60/112,850
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 60/113,296
```

```

; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 60/146,222
; PRIOR FILING DATE: July 28, 1999
; PRIOR APPLICATION NUMBER: PCT/US98/19330
; PRIOR FILING DATE: September 16, 1998
; PRIOR APPLICATION NUMBER: PCT/US98/25108
; PRIOR FILING DATE: December 1, 1998
; PRIOR APPLICATION NUMBER: 09/216,021
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 09/218,517
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 09/254,311
; PRIOR FILING DATE: March 3, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: June 22, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: September 15, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28409
; PRIOR FILING DATE: No. US20020156004A1ember 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: No. US20020156004A1ember 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28301
; PRIOR FILING DATE: December 1, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: December 16, 1999
; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: February 28, 2001
; NUMBER OF SEQ ID NOS: 120
; SEQ ID NO 7
; LENGTH: 954
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-09-944-413-7

Query Match 59.3%; Score 35; DB 9; Length 954;
Best Local Similarity 50.0%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CGXPAIXPXXTG 12
Db 411 CGADALIPVQTG 422

RESULT 9
US-09-944-403-7
; Sequence 7, Application US/09944403
; Patent No. US20020165143A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Botstein, David
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul
; APPLICANT: Grimaldi, Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Hillan, Kenneth
; APPLICANT: Kijavini, Ivar
; APPLICANT: Napier, Mary
; APPLICANT: Roy, Margaret
; APPLICANT: Tamas, Daniel
; APPLICANT: Wood, William
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P2548P1C1
; CURRENT APPLICATION NUMBER: US/09/944,403
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 09/866,028
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 60/067,411
; PRIOR FILING DATE: December 3, 1997
; PRIOR APPLICATION NUMBER: 60/069,334
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,335
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,278
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,425
; PRIOR FILING DATE: December 12, 1997
; PRIOR APPLICATION NUMBER: 60/069,696
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,694
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,702
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,870
; PRIOR FILING DATE: December 17, 1997
; PRIOR APPLICATION NUMBER: 60/069,873
; PRIOR FILING DATE: December 17, 1997
; PRIOR APPLICATION NUMBER: 60/068,017
; PRIOR FILING DATE: December 18, 1997
; PRIOR APPLICATION NUMBER: 60/070,440
; PRIOR FILING DATE: January 5, 1998
; PRIOR APPLICATION NUMBER: 60/074,086
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/074,092
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/075,945
; PRIOR FILING DATE: February 25, 1998
; PRIOR APPLICATION NUMBER: 60/112,850
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 60/113,296
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 60/146,222
; PRIOR FILING DATE: July 28, 1999
; PRIOR APPLICATION NUMBER: PCT/US98/19330
; PRIOR FILING DATE: September 16, 1998
; PRIOR APPLICATION NUMBER: PCT/US98/25108
; PRIOR FILING DATE: December 1, 1998
; PRIOR APPLICATION NUMBER: 09/216,021
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 09/218,517
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 09/254,311
; PRIOR FILING DATE: March 3, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: June 22, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: September 15, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28409
; PRIOR FILING DATE: No. US20020165143A1ember 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: No. US20020165143A1ember 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28301
; PRIOR FILING DATE: December 1, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: December 16, 1999
; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: February 11, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/04414
```



; PRIOR FILING DATE: February 22, 2000  
; PRIOR APPLICATION NUMBER: PCT/US00/05841  
; PRIOR FILING DATE: March 2, 2000  
; PRIOR APPLICATION NUMBER: PCT/US00/08439  
; PRIOR FILING DATE: March 30, 2000  
; PRIOR APPLICATION NUMBER: PCT/US00/14042  
; PRIOR FILING DATE: May 22, 2000  
; PRIOR APPLICATION NUMBER: PCT/US00/20710  
; PRIOR FILING DATE: July 28, 2000  
; PRIOR APPLICATION NUMBER: PCT/US00/32678  
; PRIOR FILING DATE: December 1, 2000  
; PRIOR APPLICATION NUMBER: PCT/US01/06520  
; PRIOR FILING DATE: February 28, 2001  
; NUMBER OF SEQ ID NOS: 120  
; SEQ ID NO 7  
; LENGTH: 954  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-09-944-403-7

Query Match 59.3%; Score 35; DB 9; Length 954;  
Best Local Similarity 50.0%; Pred. No. 42;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 12

Db 411 CGADALIPVQTG 422

RESULT 10

US-09-944-896-7

; Sequence 7, Application US/09944896

; Patent No. US20020168715A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin

; APPLICANT: Borstein, David

; APPLICANT: Eaton, Dan

; APPLICANT: Ferrara, Napoleone

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Getritsen, Mary

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul

; APPLICANT: Grunaldi, Christopher

; APPLICANT: Gurney, Austin

; APPLICANT: Hillan, Kenneth

; APPLICANT: Kljavin, Ivar

; APPLICANT: Napier, Mary

; APPLICANT: Roy, Margaret

; APPLICANT: Tomas, Daniel

; APPLICANT: Wood, William

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P2548PIC1

; CURRENT APPLICATION NUMBER: US/09/944,896

; CURRENT FILING DATE: 2001-08-31

; PRIOR APPLICATION NUMBER: 09/866,028

; PRIOR FILING DATE: 2001-05-25

; PRIOR APPLICATION NUMBER: 60/069,334

; PRIOR FILING DATE: December 11, 1997

; PRIOR APPLICATION NUMBER: 60/069,335

; PRIOR FILING DATE: December 11, 1997

; PRIOR APPLICATION NUMBER: 60/069,278

; PRIOR FILING DATE: December 11, 1997

; PRIOR APPLICATION NUMBER: 60/069,425

; PRIOR FILING DATE: December 12, 1997

; PRIOR APPLICATION NUMBER: 60/069,696

; PRIOR FILING DATE: December 16, 1997

; PRIOR APPLICATION NUMBER: 60/069,694

; PRIOR FILING DATE: December 16, 1997

; PRIOR APPLICATION NUMBER: 60/069,702

; PRIOR FILING DATE: December 16, 1997

; PRIOR APPLICATION NUMBER: 60/069,870

; PRIOR FILING DATE: December 17, 1997

; PRIOR APPLICATION NUMBER: 60/069,873  
; PRIOR FILING DATE: December 17, 1997  
; PRIOR APPLICATION NUMBER: 60/068,017  
; PRIOR FILING DATE: December 18, 1997  
; PRIOR APPLICATION NUMBER: 60/070,440  
; PRIOR FILING DATE: January 5, 1998  
; PRIOR APPLICATION NUMBER: 60/074,086  
; PRIOR FILING DATE: February 9, 1998  
; PRIOR APPLICATION NUMBER: 60/074,092  
; PRIOR FILING DATE: February 9, 1998  
; PRIOR APPLICATION NUMBER: 60/075,945  
; PRIOR FILING DATE: February 25, 1998  
; PRIOR APPLICATION NUMBER: 60/112,850  
; PRIOR FILING DATE: December 16, 1998  
; PRIOR APPLICATION NUMBER: 60/113,296  
; PRIOR FILING DATE: December 22, 1998  
; PRIOR APPLICATION NUMBER: 60/146,222  
; PRIOR FILING DATE: July 28, 1999  
; PRIOR APPLICATION NUMBER: PCT/US98/19330  
; PRIOR FILING DATE: September 16, 1998  
; PRIOR APPLICATION NUMBER: PCT/US98/25108  
; PRIOR FILING DATE: December 1, 1998  
; PRIOR APPLICATION NUMBER: 09/216,021  
; PRIOR FILING DATE: December 16, 1998  
; PRIOR APPLICATION NUMBER: 09/218,517  
; PRIOR FILING DATE: December 22, 1998  
; PRIOR APPLICATION NUMBER: 09/254,311  
; PRIOR FILING DATE: March 3, 1999  
; PRIOR APPLICATION NUMBER: PCT/US99/12252  
; PRIOR FILING DATE: June 22, 1999  
; PRIOR APPLICATION NUMBER: PCT/US99/21090  
; PRIOR FILING DATE: September 15, 1999  
; PRIOR APPLICATION NUMBER: PCT/US99/28409  
; PRIOR FILING DATE: No. US20020168715A1ember 30, 1999  
; PRIOR APPLICATION NUMBER: PCT/US99/28313  
; PRIOR FILING DATE: No. US20020168715A1ember 30, 1999  
; PRIOR APPLICATION NUMBER: PCT/US99/28301  
; PRIOR FILING DATE: December 1, 1999  
; PRIOR APPLICATION NUMBER: PCT/US99/30095  
; PRIOR FILING DATE: December 16, 1999  
; PRIOR APPLICATION NUMBER: PCT/US00/03565  
; PRIOR FILING DATE: February 11, 2000  
; PRIOR APPLICATION NUMBER: PCT/US00/04414  
; PRIOR FILING DATE: February 22, 2000  
; PRIOR APPLICATION NUMBER: PCT/US00/05841  
; PRIOR FILING DATE: March 2, 2000  
; PRIOR APPLICATION NUMBER: PCT/US00/08439  
; PRIOR FILING DATE: March 30, 2000  
; PRIOR APPLICATION NUMBER: PCT/US00/14042  
; PRIOR FILING DATE: May 22, 2000  
; PRIOR APPLICATION NUMBER: PCT/US00/20710  
; PRIOR FILING DATE: July 28, 2000  
; PRIOR APPLICATION NUMBER: PCT/US00/32678  
; PRIOR FILING DATE: December 1, 2000  
; PRIOR APPLICATION NUMBER: PCT/US01/06520  
; PRIOR FILING DATE: February 28, 2001  
; NUMBER OF SEQ ID NOS: 120  
; SEQ ID NO 7  
; LENGTH: 954  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-09-944-896-7

Query Match 59.3%; Score 35; DB 9; Length 954;  
Best Local Similarity 50.0%; Pred. No. 42;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 12

Db 411 CGADALIPVQTG 422

RESULT 11

US-09-944-944-7  
; Sequence 7, Application US/09944944  
; Patent No. US20020173463A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin  
; APPLICANT: Botstein, David  
; APPLICANT: Eaton, Dan  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul  
; APPLICANT: Grimaldi, Christopher  
; APPLICANT: Gurney, Austin  
; APPLICANT: Hillan, Kenneth  
; APPLICANT: Kijavin, Ivar  
; APPLICANT: Napier, Mary  
; APPLICANT: Roy, Margaret  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Wood, William  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P2548P1C1  
; CURRENT FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: 09/866,028  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: 60/067,411  
; PRIOR FILING DATE: December 3, 1997  
; PRIOR APPLICATION NUMBER: 60/069,334  
; PRIOR FILING DATE: December 11, 1997  
; PRIOR APPLICATION NUMBER: 60/069,335  
; PRIOR FILING DATE: December 11, 1997  
; PRIOR APPLICATION NUMBER: 60/069,278  
; PRIOR FILING DATE: December 11, 1997  
; PRIOR APPLICATION NUMBER: 60/069,425  
; PRIOR FILING DATE: December 12, 1997  
; PRIOR APPLICATION NUMBER: 60/069,696  
; PRIOR FILING DATE: December 16, 1997  
; PRIOR APPLICATION NUMBER: 60/069,694  
; PRIOR FILING DATE: December 16, 1997  
; PRIOR APPLICATION NUMBER: 60/069,702  
; PRIOR FILING DATE: December 16, 1997  
; PRIOR APPLICATION NUMBER: 60/069,870  
; PRIOR FILING DATE: December 17, 1997  
; PRIOR APPLICATION NUMBER: 60/069,873  
; PRIOR FILING DATE: December 17, 1997  
; PRIOR APPLICATION NUMBER: 60/068,017  
; PRIOR FILING DATE: December 18, 1997  
; PRIOR APPLICATION NUMBER: 60/070,440  
; PRIOR FILING DATE: January 5, 1998  
; PRIOR APPLICATION NUMBER: 60/074,086  
; PRIOR FILING DATE: February 9, 1998  
; PRIOR APPLICATION NUMBER: 60/074,092  
; PRIOR FILING DATE: February 9, 1998  
; PRIOR APPLICATION NUMBER: 60/075,945  
; PRIOR FILING DATE: February 25, 1998  
; PRIOR APPLICATION NUMBER: 60/112,850  
; PRIOR FILING DATE: December 16, 1998  
; PRIOR APPLICATION NUMBER: 60/113,236  
; PRIOR FILING DATE: December 22, 1998  
; PRIOR APPLICATION NUMBER: 60/146,222  
; PRIOR FILING DATE: July 28, 1999  
; PRIOR APPLICATION NUMBER: PCT/US98/19330  
; PRIOR FILING DATE: September 16, 1998  
; PRIOR APPLICATION NUMBER: PCT/US98/25108  
; PRIOR FILING DATE: December 1, 1998  
; PRIOR APPLICATION NUMBER: 09/216,021  
; PRIOR FILING DATE: December 16, 1998  
; PRIOR APPLICATION NUMBER: 09/218,517  
; PRIOR FILING DATE: December 22, 1998  
; PRIOR APPLICATION NUMBER: 09/254,311  
; PRIOR FILING DATE: March 3, 1999

; PRIOR APPLICATION NUMBER: PCT/US99/12252  
; PRIOR FILING DATE: June 22, 1999  
; PRIOR APPLICATION NUMBER: PCT/US99/21090  
; PRIOR FILING DATE: September 15, 1999  
; PRIOR APPLICATION NUMBER: PCT/US99/28409  
; PRIOR FILING DATE: No. US20020173463A1ember 30, 1999  
; PRIOR APPLICATION NUMBER: PCT/US99/28313  
; PRIOR FILING DATE: No. US20020173463A1ember 30, 1999  
; PRIOR APPLICATION NUMBER: PCT/US99/28301  
; PRIOR FILING DATE: December 1, 1999  
; PRIOR APPLICATION NUMBER: PCT/US99/30095  
; PRIOR FILING DATE: December 16, 1999  
; PRIOR APPLICATION NUMBER: PCT/US00/03565  
; PRIOR FILING DATE: February 11, 2000  
; PRIOR APPLICATION NUMBER: PCT/US00/04414  
; PRIOR FILING DATE: February 22, 2000  
; PRIOR APPLICATION NUMBER: PCT/US00/05841  
; PRIOR FILING DATE: March 2, 2000  
; PRIOR APPLICATION NUMBER: PCT/US00/08439  
; PRIOR FILING DATE: March 30, 2000  
; PRIOR APPLICATION NUMBER: PCT/US00/14042  
; PRIOR FILING DATE: May 22, 2000  
; PRIOR APPLICATION NUMBER: PCT/US00/20710  
; PRIOR FILING DATE: July 28, 2000  
; PRIOR APPLICATION NUMBER: PCT/US00/32678  
; PRIOR FILING DATE: December 1, 2000  
; PRIOR APPLICATION NUMBER: PCT/US01/06520  
; PRIOR FILING DATE: February 28, 2001  
; NUMBER OF SEQ ID NOS: 120  
; SEQ ID NO 7  
; LENGTH: 954  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-09-944-944-7  
  
Query Match 59.3%; Score 35; DB 9; Length 954;  
Best Local Similarity 50.0%; Pred. No. 42;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
  
QY 1 CGXPATXPPXKT 12  
Db 411 CGDALIPVQTG 422  
  
RESULT 12  
US-09-944-907-7  
; Sequence 7, Application US/09944907  
; Publication No. US20020198147A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin  
; APPLICANT: Botstein, David  
; APPLICANT: Baton, Dan  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul  
; APPLICANT: Grimaldi, Christopher  
; APPLICANT: Gurney, Austin  
; APPLICANT: Hillan, Kenneth  
; APPLICANT: Kijavin, Ivar  
; APPLICANT: Napier, Mary  
; APPLICANT: Roy, Margaret  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Wood, William  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P2548P1C1  
; CURRENT APPLICATION NUMBER: US/09/944,907  
; CURRENT FILING DATE: 2001-08-31  
; PRIOR APPLICATION NUMBER: 09/866,028  
; PRIOR FILING DATE: 2001-05-25  
; NUMBER OF SEQ ID NOS: 120

```

; SEQ ID NO 7
; LENGTH: 954
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-944-907-7

Query Match          59.3%; Score 35; DB 9; Length 954;
Best Local Similarity 50.0%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGXPALPXXXTG 12
    |||:|:|
Db 411 CGADALIPVQTG 422

RESULT 13
US-09-944-929-7
; Sequence 7, Application US/09944929
; Publication No. US20020197612A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Botstein, David
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul
; APPLICANT: Grimaldi, Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Hillan, Kenneth
; APPLICANT: Kljavin, Ivar
; APPLICANT: Napier, Mary
; APPLICANT: Roy, Margaret
; APPLICANT: Tumas, Daniel
; APPLICANT: Wood, William
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P2548P1C1
; CURRENT APPLICATION NUMBER: US/09/944,929
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: 09/866,028
; NUMBER OF SEQ ID NOS: 120
; SEQ ID NO 7
; LENGTH: 954
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-944-929-7

Query Match          59.3%; Score 35; DB 9; Length 954;
Best Local Similarity 50.0%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGXPALPXXXTG 12
    |||:|:|
Db 411 CGADALIPVQTG 422

RESULT 14
US-09-866-028-7
; Sequence 7, Application US/09866028
; Patent No. US20020058309A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Botstein, David
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul
; APPLICANT: Grimaldi, Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Hillan, Kenneth
; APPLICANT: Kljavin, Ivar
; APPLICANT: Napier, Mary
; APPLICANT: Roy, Margaret
; APPLICANT: Tumas, Daniel
; APPLICANT: Wood, William
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P2548P1C1
; CURRENT APPLICATION NUMBER: US/09/944,929
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: 09/866,028
; NUMBER OF SEQ ID NOS: 120
; SEQ ID NO 7
; LENGTH: 954
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-944-929-7

Query Match          59.3%; Score 35; DB 9; Length 954;
Best Local Similarity 50.0%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGXPALPXXXTG 12
    |||:|:|
Db 411 CGADALIPVQTG 422

RESULT 15
US-09-944-449-7
; Sequence 7, Application US/09944449
; Patent No. US20020102647A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Botstein, David
; APPLICANT: Baton, Dan
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul
; APPLICANT: Grimaldi, Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Hillan, Kenneth
; APPLICANT: Kljavin, Ivar
; APPLICANT: Napier, Mary
; APPLICANT: Roy, Margaret
; APPLICANT: Tumas, Daniel
; APPLICANT: Wood, William
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P2548P1C1
; CURRENT APPLICATION NUMBER: US/09/944,449
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 09/866,028
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 60/067,411
; PRIOR FILING DATE: December 3, 1997
; PRIOR APPLICATION NUMBER: 60/069,334
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,335
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,278
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,425
; PRIOR FILING DATE: December 12, 1997
; PRIOR APPLICATION NUMBER: 60/069,696
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,634
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,702
; PRIOR FILING DATE: December 16, 1997

```

us-10-036-371-3.rapb

Wed Feb 12 11:59:35 2003

Search completed: February 12, 2003, 10:31:05  
Job time : 2.9791 secs

;; PRIOR APPLICATION NUMBER: 60/069,870  
;; PRIOR FILING DATE: December 17, 1997  
;; PRIOR APPLICATION NUMBER: 60/069,873  
;; PRIOR FILING DATE: December 17, 1997  
;; PRIOR APPLICATION NUMBER: 60/068,017  
;; PRIOR FILING DATE: December 18, 1997  
;; PRIOR APPLICATION NUMBER: 60/070,440  
;; PRIOR FILING DATE: January 5, 1998  
;; PRIOR APPLICATION NUMBER: 60/074,086  
;; PRIOR FILING DATE: February 9, 1998  
;; PRIOR APPLICATION NUMBER: 60/074,092  
;; PRIOR FILING DATE: February 9, 1998  
;; PRIOR APPLICATION NUMBER: 60/075,945  
;; PRIOR FILING DATE: February 25, 1998  
;; PRIOR APPLICATION NUMBER: 60/112,850  
;; PRIOR FILING DATE: December 16, 1998  
;; PRIOR APPLICATION NUMBER: 60/113,296  
;; PRIOR FILING DATE: December 22, 1998  
;; PRIOR APPLICATION NUMBER: 60/146,222  
;; PRIOR FILING DATE: July 28, 1999  
;; PRIOR APPLICATION NUMBER: PCT/US98/19330  
;; PRIOR FILING DATE: September 16, 1998  
;; PRIOR APPLICATION NUMBER: PCT/US98/25108  
;; PRIOR FILING DATE: December 1, 1998  
;; PRIOR APPLICATION NUMBER: 09/216,021  
;; PRIOR FILING DATE: December 16, 1998  
;; PRIOR APPLICATION NUMBER: 09/218,517  
;; PRIOR FILING DATE: December 22, 1998  
;; PRIOR APPLICATION NUMBER: 09/254,311  
;; PRIOR FILING DATE: March 3, 1999  
;; PRIOR APPLICATION NUMBER: PCT/US99/12252  
;; PRIOR FILING DATE: June 22, 1999  
;; PRIOR APPLICATION NUMBER: PCT/US99/21090  
;; PRIOR FILING DATE: September 15, 1999  
;; PRIOR APPLICATION NUMBER: PCT/US99/28409  
;; PRIOR FILING DATE: No. US20020102647A1ember 30, 1999  
;; PRIOR APPLICATION NUMBER: PCT/US99/28313  
;; PRIOR FILING DATE: No. US20020102647A1ember 30, 1999  
;; PRIOR APPLICATION NUMBER: PCT/US99/28301  
;; PRIOR FILING DATE: December 1, 1999  
;; PRIOR APPLICATION NUMBER: PCT/US99/30095  
;; PRIOR FILING DATE: December 16, 1999  
;; PRIOR APPLICATION NUMBER: PCT/US00/03565  
;; PRIOR FILING DATE: February 11, 2000  
;; PRIOR APPLICATION NUMBER: PCT/US00/04414  
;; PRIOR FILING DATE: February 22, 2000  
;; PRIOR APPLICATION NUMBER: PCT/US00/05841  
;; PRIOR FILING DATE: March 2, 2000  
;; PRIOR APPLICATION NUMBER: PCT/US00/08439  
;; PRIOR FILING DATE: March 30, 2000  
;; PRIOR APPLICATION NUMBER: PCT/US00/14042  
;; PRIOR FILING DATE: May 22, 2000  
;; PRIOR APPLICATION NUMBER: PCT/US00/20710  
;; PRIOR FILING DATE: July 28, 2000  
;; PRIOR APPLICATION NUMBER: PCT/US00/32678  
;; PRIOR FILING DATE: December 1, 2000  
;; PRIOR APPLICATION NUMBER: PCT/US01/06520  
;; PRIOR FILING DATE: February 28, 2001  
;; NUMBER OF SEQ ID NOS: 120  
;; SEQ ID NO 7  
;; LENGTH: 954  
;; TYPE: PRT  
;; ORGANISM: Homo Sapien  
US-09-944-449-7

Query Match 59.3%; Score 35; DB 10; Length 954;  
Best Local Similarity 50.0%; Pred. No. 42;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGXPAIXPXXTG 12

Db 411 CGADALIPVQTG 422

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:03:55 ; Search time 13.4328 Seconds  
(without alignment)  
198.395 Million cell updates/sec

Title: US-10-036-371-4  
Perfect score: 100  
Sequence: 1 IINGEEAVPHXWQVLSQX 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002.\*  
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	95	95.0	20	22	AAB31576 Amino terminal seq
2	85	85.0	20	22	AAB31578 Amino terminal seq
3	85	85.0	31	17	AAW02588 PX, porcine pancre
4	85	85.0	242	7	AAP60061 Sequence of human
5	85	85.0	245	21	AAV99596 Bovine chymotrypsi
6	85	85.0	253	13	AAK29621 Porcine pancreatic
7	82	82.0	20	20	AAV33344 Bovine chymotrypsi
8	82	82.0	20	21	AAV93935 N-terminal of a bo
9	82	82.0	20	22	AAE07941 N-terminal of bovi
10	82	82.0	146	21	AAK54191 Human pancreatic c

11	82	82.0	231	22	AAB98504 Human chymotrypsin
12	82	82.0	263	23	AAU82738 Amino acid sequenc
13	80	80.0	242	7	AAP60060 Sequence of human
14	80	80.0	264	21	AAK11711 Mouse serine prote
15	80	80.0	270	8	AAP70759 Human pancreas ela
16	80	80.0	270	23	AAU87891 Human pancreatic t
17	79	79.0	242	13	AAK20510 Bile acid secretio
18	78	78.0	29	16	AAK84270 Serum calcium cont
19	78	78.0	268	17	AAK90682 Rat caldesmon cont
20	78	78.0	273	21	AAK54243 Human pancreatic c
21	77	77.0	268	17	AAK90683 Human pancreatic c
22	77	77.0	268	17	AAK88481 Human caldesmon c
23	77	77.0	268	19	AAK40530 Human elastase IV
24	77	77.0	268	20	AAK89410 Human elastase hom
25	77	77.0	268	21	AAK51839 Human elastase IV
26	76	76.0	123	20	AAK11850 Human elastase IV
27	74	74.0	195	20	AAK11649 Human 5' EST seque
28	74	74.0	252	7	AAP60058 Sequence of human
29	74	74.0	269	7	AAP60062 Human elastase II.
30	74	74.0	269	7	AAP61723 Human elastase I.
31	74	74.0	269	21	AAK54124 Human pancreatic c
32	74	74.0	269	22	ABG20513 Porcine elastase I
33	74	74.0	279	22	ABG20513 Pig pancreas elast
34	71	71.0	269	7	AAP61724 Human serine prote
35	71	71.0	269	8	AAP70758 Novel human diago
36	70	70.0	192	21	AAK54077 Amino acid sequenc
37	70	70.0	264	21	AAK11710 Novel human protei
38	70	70.0	913	22	ABG24246 Human pancreatic c
39	70	70.0	970	23	AAU82743 Novel human diago
40	70	70.0	1031	23	ABG60993 Amino acid sequenc
41	70	70.0	1576	22	ABG07870 Novel human protei
42	70	70.0	1576	22	ABG10218 Novel human diago
43	70	70.0	1576	22	ABG14588 Novel human diago
44	70	70.0	1576	22	ABG19687 Novel human diago
45	69	69.0	113	21	AAK54419 Human pancreatic c

#### ALIGNMENTS

RESULT 1  
AAB31576  
ID AAB31576 standard; peptide; 20 AA.  
XX AAB31576;  
AC AAB31576;  
XX AAB31576;  
DT 20-APR-2001 (first entry)  
XX Amino terminal sequence of cod trypsin B chain.

XX Amino terminal sequence of cod trypsin B chain.  
XX Fish; serine proteinase; pain; acute inflammation; chronic inflammation;  
XX arthritis; inflamed joint; bursitis; osteoarthritis; septic arthritis;  
XX rheumatoid arthritis; juvenile rheumatoid arthritis; fibromyalgia;  
XX systemic lupus erythematosus; phlebitis; tendinitis; rash; psoriasis;  
XX acne; eczema; facial seborethic eczema; foreskin infection;  
XX athlete's foot; fistulae infection; ulcer; navel infection;  
XX scar; keloid; boil; wart; allergic itch; hemorrhoid; wrinkle;  
XX fungal infection; autoimmune disease.  
XX Gadus callarias.  
XX OS

XX Key Location/Qualifiers  
XX Misc-difference 11 /label= Ser, Thr  
XX Misc-difference 13 /label= Ser, Pro, Tyr  
XX Misc-difference 20 /label= Asp, Gln  
XX WO200078332-A2.  
XX N-terminal of a bo  
XX 28-DEC-2000.

```

PF 15-JUN-2000; 2000WO-IS000005.
XX
XX
PR 18-JUN-1999; 99IS-0005086.
XX
XX
PA (BJAR/) BJARNASON J B.
XX
XX
PI Bjarnason JB;
XX
XX
XX WPI; 2001-091493/10.
XX
XX
XX Fish serine proteinase, useful as a cosmetic, medicament for treating
XX PT eczema, psoriasis, arthritis, and in the manufacture of the medicament
XX PT for treating, preventing pathogenic diseases involving receptor
XX PT mediated binding -
XX
XX PS Disclosure; Page 5; 38pp; English.
XX
XX CC The specification describes a fish serine proteinase. The proteinases
XX CC are useful as medicaments, for treating and preventing a disease in a
XX CC human or an animal such as pain, acute inflammation, chronic
XX CC inflammation, arthritis, juvenile rheumatoid arthritis, septic arthritis,
XX CC rheumatoid arthritis, systemic lupus erythematosus, phlebitis, tendinitis, rash,
XX CC fibromyalgia, systemic lupus erythematosus, phlebitis, tendinitis, rash,
XX CC psoriasis, acne, eczema, facial seborrheic eczema, eczema of the hands,
XX CC face or neck, foreskin infections, athlete's foot, fistulae infections,
XX CC infected topical ulcers, navel infections in newborns, wrinkles, scars,
XX CC keloids, boils, warts and allergic itch, hemorrhoids, wounds, wound
XX CC infections, wounds from burns, fungal infections and immunological and
XX CC autoimmune diseases. They are also useful for removing dead or peeling
XX CC skin from otherwise healthy skin, and for treating or preventing a
XX CC disease in which pathogenesis is caused by bacteria, virus, fungus,
XX CC parasite or a protozoan or a receptor mediated binding is involved.
XX CC The present sequence represents the amino terminal of cod trypsin B
XX CC chain.
XX
XX SQ Sequence 20 AA;
XX
XX Query Match 95.0%; Score 95; DB 22; Length 20;
XX Best Local Similarity 100.0%; Pred. No. 3.4e-09;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 IVNGEEAVPHXMXWQVSLQ 19
XX |||||
XX Db 1 IVNGEEAVPHXMXWQVSLQ 19
XX
XX RESULT 2
XX AAB31578
XX ID AAB31578 standard; peptide; 20 AA.
XX
XX AC AAB31578;
XX
XX DT 20-APR-2001 (first entry)
XX
XX DE Amino terminal sequence of bovine trypsin B chain.
XX
XX KW Fish; serine proteinase; pain; acute inflammation; chronic inflammation;
XX KW arthritis; inflamed joint; bursitis; osteoarthritis; septic arthritis;
XX KW rheumatoid arthritis; juvenile rheumatoid arthritis; fibromyalgia;
XX KW systemic lupus erythematosus; phlebitis; tendinitis; rash; psoriasis;
XX KW acne; eczema; facial seborrheic eczema; foreskin infection;
XX KW athlete's foot; fistulae infection; ulcer; navel infection; wrinkle;
XX KW scar; keloid; boil; wart; allergic itch; hemorrhoid; wound;
XX KW fungal infection; autoimmune disease.
XX
XX OS Bos sp.
XX
XX PN WO200078332-A2.
XX
XX PD 28-DEC-2000.
XX
XX PF 15-JUN-2000; 2000WO-IS000005.
XX

```

---

```

PR 18-JUN-1999; 99IS-0005086.
XX
XX
PA (BJAR/) BJARNASON J B.
XX
XX
PI Bjarnason JB;
XX
XX
XX WPI; 2001-091493/10.
XX
XX
XX Fish serine proteinase, useful as a cosmetic, medicament for treating
XX PT eczema, psoriasis, arthritis, and in the manufacture of the medicament
XX PT for treating, preventing pathogenic diseases involving receptor
XX PT mediated binding -
XX
XX PS Disclosure; Page 5; 38pp; English.
XX
XX CC The specification describes a fish serine proteinase. The proteinases
XX CC are useful as medicaments, for treating and preventing a disease in a
XX CC human or an animal such as pain, acute inflammation, chronic
XX CC inflammation, arthritis, juvenile rheumatoid arthritis, septic arthritis,
XX CC rheumatoid arthritis, juvenile lupus erythematosus, phlebitis, tendinitis, rash,
XX CC fibromyalgia, systemic lupus erythematosus, phlebitis, tendinitis, rash,
XX CC psoriasis, acne, eczema, facial seborrheic eczema, eczema of the hands,
XX CC face or neck, foreskin infections, athlete's foot, fistulae infections,
XX CC infected topical ulcers, navel infections in newborns, wrinkles, scars,
XX CC keloids, boils, warts and allergic itch, hemorrhoids, wounds, wound
XX CC infections, wounds from burns, fungal infections and immunological and
XX CC autoimmune diseases. They are also useful for removing dead or peeling
XX CC skin from otherwise healthy skin, and for treating or preventing a
XX CC disease in which pathogenesis is caused by bacteria, virus, fungus,
XX CC parasite or a protozoan or a receptor mediated binding is involved.
XX CC The present sequence represents the amino terminal of bovine trypsin B
XX CC chain.
XX
XX SQ Sequence 20 AA;
XX
XX Query Match 85.0%; Score 85; DB 22; Length 20;
XX Best Local Similarity 84.2%; Pred. No. 1.6e-07;
XX Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 1 IVNGEEAVPHXMXWQVSLQ 19
XX |||||
XX Db 1 IVNGEEAVPGSWPQVSLQ 19
XX
XX RESULT 3
XX AAW02588
XX ID AAW02588 standard; peptide; 31 AA.
XX
XX AC AAW02588;
XX
XX DT 12-FEB-1997 (first entry)
XX
XX DE PX, porcine pancreas-derived factor N-terminal peptide.
XX
XX KW Pig; bone; calcium; resorption inhibition; formation; osteoporosis;
XX KW Paget's disease; fracture repair; bone defect; osteopetrosis;
XX KW metastasis; osteoblast; osteoclast.
XX
XX OS Sus scrofa.
XX
XX PN WO9619501-A1.
XX
XX PD 27-JUN-1996.
XX
XX PF 19-DEC-1995; 95WO-US16826.
XX
XX PR 20-DEC-1994; 94US-0363092.
XX
XX PA (OSTE-) OSTEOA INC.
XX
XX PI Izbicka E, Mundy GR, Yoneda T;
XX
XX MPI; 1996-309523/31.
XX

```



```

RESULT 6
AAK29621
ID AAR29621 standard; Protein; 253 AA.
XX
XX AAR29621;
AC
XX
XX 30-APR-1993 (first entry)
DT
XX
XX Porcine pancreatic elastase III.
DE
XX Pig; swine; bile acid; gall bladder; bile acid secretion promoters;
KW liver function improvers; ss.
KW
XX
XX Sus scrofa domestica.
OS
XX JP04325090-A.
PN
XX
XX 13-NOV-1992.
PD
XX 23-APR-1991; 91JP-0092069.
PF
XX 23-APR-1991; 91JP-0092069.
PR
XX (SANY ) SANKYO CO LTD.
PA
XX WPI; 1992-428826/52.
XX
XX N-PSDB; AAQ31724.
DR
XX
XX Pig pancreas elastase protein - used in bile acid secretion
PT promoters and liver function improvers
XX
XX Claim 1; Page 8; 10pp; Japanese.
PS
XX
XX This sequence represents porcine pancreatic elastase III. It was
CC expressed in E. coli YAZ1 using the expression vector pELE001.
CC It may be used in bile acid secretion promoters and liver function
CC improvers.
XX
XX Sequence 253 AA;
SQ
Query Match 85.0%; Score 85; DB 13; Length 253;
Best Local Similarity 73.7%; Pred. No. 2.8e-06;
Matches 14; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSLQ 19
:|||||:|||||
Db 12 VVNGEDAVPSWQVSLQ 30

RESULT 7
AAV33344
ID AAV33344 standard; peptide; 20 AA.
XX
XX AAV33344;
AC
XX
XX 29-NOV-1999 (first entry)
DT
XX
XX Bovine chymotrypsin N-terminal fragment.
DE
XX
XX Multifunctional enzyme; treatment; acne; eczema; chymotrypsin; trypsin;
KW collagenase; elastase; exopeptidase; cell surface receptor; anti-acne;
KW antisborrheic; dermatological; anti-eczema; adhesion molecule; ICAM-1;
KW CD54; ICAM-2; VCAM-1; CD4; CD8; CD28; CD29D; CD31; CD44; CD49; CD62L;
KW CD102; GMI ceramide; krill; bovine.
XX
XX Bos taurus.
OS
XX
XX US5958406-A.
PN
XX
XX 28-SEP-1999.
PD
XX
XX 08-FEB-1996; 96US-0600273.
PF

```

```

XX 22-NOV-1994; 94US-0388501.
PR 08-FEB-1995; 95US-0385540.
PR 07-JUN-1995; 95US-0486820.
XX
XX (PHAI-) PHAIRSON MEDICAL INC.
PA
XX Lindblom R, Kay J, Franklin RL, De Faire JR;
PI
XX WPI; 1999-561004/47.
XX
XX Treating acne and eczema using a krill-derived multifunctional enzyme -
PT Disclosure; Column 21-22; 42pp; English.
XX
XX This invention describes a novel method for treating acne and eczema
CC using a krill-derived multifunctional enzyme (I) which comprises 2 or
CC more of the activities of chymotrypsin, trypsin, collagenase, elastase
CC or exopeptidase and is reactive with cell surface receptors such as
CC proteins or glycoproteins. The product of the invention have
CC antisborrheic, anti-acne, dermatological and anti-eczema activity.
CC (I) removes or inactivates cell surface receptors (proteins and
CC glycoproteins) and adhesion molecules such as ICAM-1 (i.e. CD54), ICAM-2,
CC VCAM-1, CD4, CD8, CD28, CD29D, CD31, CD44, CD49, CD62L CD102 and the
CC asialo GMI ceramide. This sequence represents the N-terminal fragment of
CC bovine chymotrypsin which is used to describe the method of the
CC invention.
XX
XX Sequence 20 AA;
SQ
Query Match 82.0%; Score 82; DB 20; Length 20;
Best Local Similarity 78.9%; Pred. No. 5.2e-07;
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSLQ 19
:|||||:|||||
Db 1 IVNGEDAVPGSWQVSLQ 19

RESULT 8
AAV93935
ID AAV93935 standard; peptide; 20 AA.
XX
XX AAV93935;
AC
XX
XX 03-OCT-2000 (first entry)
DT
XX
XX N-terminal of a bovine chymotrypsin hydrolase enzyme.
DE
XX
XX Transplantation rejection; hydrolase; graft versus host disease;
KW cell surface adhesion molecule; immune reaction; inflammation; shock;
KW tumour metastasis; autoimmune disease;
KW Krill derived multifunctional enzyme.
XX
XX Bos sp.
OS
XX
XX WC2000038708-A1.
PN
XX
XX 06-JUL-2000.
PD
XX
XX 23-DEC-1999; 99WO-US30818.
PF
XX 24-DEC-1998; 98US-0114147.
PR
XX (PHAI-) PHAIRSON MEDICAL INC.
PA
XX
XX Franklin RL, St Pierre Y;
PI
XX WPI; 2000-452301/39.
XX
XX Preventing or ameliorating transplantation rejection reactions using
PT hydrolase enzymes -
XX

```



PS Disclosure; Page 26; 66pp; English.

XX The specification describes a method for preventing or ameliorating

CC transplantation rejection reactions for transplantation of immune cells

CC or other tissues. The method comprises treating a source of immune cells

CC with a hydrolase or hydrolase mixture and administering the treated

CC immune cells to a recipient animal. The hydrolase especially has a

CC preference for removing, destroying, inactivating or disabling at least

CC one of CD4, CD8, CD25, ICAM-1 (CD54), CD152, an integrin, CD154,

CC CD40 and CD80 in contrast to removing, destroying, inactivating or

CC disabling TCR. The methods are useful for preventing graft versus host

CC disease by using hydrolase enzymes to remove the cell surface adhesion

CC molecules which are involved in triggering the immune reactions involved

CC in the diseases. The methods are used for treating or preventing

CC cell-cell or cell-virus adhesion syndrome comprising inflammation, shock,

CC tumour metastases, autoimmune disease, transplantation rejection

CC reactions or microbial infections. The present sequence represents the

CC N-terminal of a hydrolase, which may be used in the method of the

CC invention.

XX

SQ Sequence 20 AA;

Query Match 82.0%; Score 82; DB 21; Length 20;

Best Local Similarity 78.9%; Pred. No. 5.2e-07;

Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSLQ 19

|||||:|||||

Db 1 IVNGEDAVPGSWPQVSLQ 19

|||||:|||||

RESULT 9

AAE07941

ID AAE07941 standard; peptide; 20 AA.

AC AAE07941;

XX

DT 01-NOV-2001 (first entry)

XX

DE N-terminal of bovine serine multifunctional hydrolase.

XX

KW Multifunctional hydrolase; rejection reaction; extra-corporeal;

KW therapy; graft versus host disease; transplantation rejection;

KW autoimmune disease; microbial infection; immune disorder; cytostatic;

KW cystic fibrosis; chronic obstructive pulmonary disease; COPD;

KW atherosclerosis; cancer; asthma; septic shock; toxic shock syndrome;

KW conjunctivitis; reperfusion injury; pain; immunosuppressive;

KW antibacterial; vasotropic; bovine.

XX

OS Bos sp.

XX

PN US6232088-B1.

XX

PD 15-MAY-2001.

XX

PF 24-DEC-1998; 98US-0220731.

XX

PR 08-FEB-1995; 95US-0385540.

PR 07-JUN-1995; 95US-0486820.

PR 08-FEB-1996; 96US-0600273.

XX

PA (PHAI-) PHAIRSON MEDICAL INC.

XX

PI Franklin R, St Pierre Y;

XX

DR WPI; 2001-450051/48.

XX

PT Ameliorating, preventing or treating immune rejection reactions, such

PT as graft versus host disease, autoimmune disease, asthma, cancer, by

PT extra-corporeally treating donor tissue with hydrolase such as a

PT protease.

XX

PS Disclosure; Column 15-16; 27pp; English.

XX The present invention relates to a method for ameliorating

CC transplantation rejection reaction. The method comprises treating

CC extra-corporeally donor tissue or donor source of immune cells with

CC a rejection ameliorating effective amount of a hydrolase enzyme. The

CC method is useful for ameliorating, treating or preventing immune

CC rejection reactions such as graft versus host diseases, organ or tissue

CC transplantation rejection, autoimmune disease and associated conditions,

CC microbial infection, immune disorder, cystic fibrosis, chronic

CC obstructive pulmonary disease (COPD), atherosclerosis, cancer, asthma,

CC septic shock, toxic shock syndrome, conjunctivitis, reperfusion injury

CC and pain in humans. The present sequence is the N-terminal peptide

CC of bovine serine multifunctional hydrolase, used in the

CC exemplification of the invention.

XX

SQ Sequence 20 AA;

Query Match 82.0%; Score 82; DB 22; Length 20;

Best Local Similarity 78.9%; Pred. No. 5.2e-07;

Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSLQ 19

|||||:|||||

Db 1 IVNGEDAVPGSWPQVSLQ 19

|||||:|||||

RESULT 10

AAE07941

ID AAE07941 standard; Protein; 146 AA.

XX

AC AAE07941;

XX

DT 09-MAR-2001 (first entry)

XX

DE Human pancreatic cancer antigen protein sequence SEQ ID NO:643.

XX

KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;

KW detection; diagnosis; identification; cytostatic; neuroprotective;

KW neotropic; immunomodulatory; relaxant; contraceptive; gynaecological;

KW antiinflammatory; cardiant; gene therapy; chromosome mapping;

KW linkage analysis; tissue identification; tissue typing; forensic;

KW neural; immune system; muscular; reproductive; gastrointestinal;

KW pulmonary; cardiovascular; renal; proliferative.

XX

OS Homo sapiens.

XX

PN WO200055320-A1.

XX

PD 21-SEP-2000.

XX

PF 08-MAR-2000; 2000WO-US05989.

XX

PR 12-MAR-1999; 99US-0124270.

XX

XX (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Ruben SM;

XX

DR WPI; 2000-579444/54.

DR N-PSDB; AAC98956.

XX

PT New nucleic acid that is a pancreatic cancer antigen for preventing,

PT treating, or ameliorating a medical condition, particular pancreatic

PT cancer, or for use in assays for diagnosing a pathological condition -

XX

PS Claim 11; Page 1081; 1379pp; English.

XX

CC AAC98773 to AAC99231 encode the human pancreatic cancer associated

CC proteins, called pancreatic cancer antigens, given in AAB54068 to

CC AAB54466. The human pancreatic cancer antigens have cytostatic,

CC neuroprotective, neotropic, immunomodulatory, relaxant, contraceptive,

CC gynaecological, cardiant and antiinflammatory activities, and can be used

CC in gene therapy. The polynucleotide and proteins can be used for

CC preventing, treating, or ameliorating a medical condition or in assays  
CC for diagnosing a pathological condition or a susceptibility to one in a  
CC subject. Binding partners to the proteins and the activity of the  
CC protein can be identified. The pancreatic cancer antigens can be used to  
CC detect, treat or prevent pancreatic disorders, especially cancer.  
CC Agonists and antagonists to the antigens can be screened for. The  
CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
CC acid hybridisation probes that can be used in chromosome mapping, linkage  
CC analysis, tissue identification and/or typing and a variety of forensic  
CC and diagnostic methods. The proteins can be used to generate antibodies  
CC which are used to purify, detect and target the polypeptides, including  
CC both in vivo and in vitro diagnostic and therapeutic methods. The  
CC proteins can be used to treat or prevent neural immune system, muscular,  
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
CC sequences used in the exemplification of the present invention.

XX Sequence 146 AA;  
SQ

Query Match 82.0%; Score 82; DB 21; Length 146;  
Best Local Similarity 78.9%; Pred. No. 4.9e-06;  
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEAVPHXWQVSLQ 19  
Db 38 IVNGEDAVPGSPWQVSLQ 56

RESULT 11  
AAB98504  
ID AAB98504 standard; Protein; 231 AA.  
XX  
AC AAB98504;  
XX  
XX 03-AUG-2001 (first entry)  
XX Human chymotrypsin serine protease catalytic domain.  
DE  
XX Human; TADG-15; cytostatic; vaccine; ovarian tumour; cancer;  
KW tumour antigen-derived gene 15; serine protease; chymotrypsin.  
KW  
XX Homo sapiens.  
OS  
XX WO200129056-A1.  
PN  
XX 26-APR-2001.  
PD  
XX 20-OCT-2000; 2000WO-US29095.  
PF  
XX 20-OCT-1999; 99US-0421213.  
PR  
XX (UYAR-) UNIV ARKANSAS.  
PA  
XX O'Brien TJ, Tanimoto H;  
PI  
XX WPI; 2001-381031/40.  
DR  
XX Novel extracellular serine protease, termed tumor antigen-derived gene  
PT 15 protein overexpressed in carcinomas and DNA encoding it, for  
PT diagnosis, treatment, prevention of cancer, particularly breast,  
PT ovarian cancer -  
PT  
XX  
PS Example 10; Fig 1; 130pp; English.  
XX  
CC The present invention relates to human tumour antigen-derived gene 15  
CC (TADG-15) protein and coding sequence (see AAB93601 and AAB98500).  
CC TADG-15 is an extracellular serine protease. It was found that TADG-15 is  
CC over-expressed in ovarian tumours. TADG-15 protein or its fragments of  
CC 9-20 residues that lack TADG-15 protease activity are useful for  
CC vaccinating an individual against TADG-15, having, suspected of having or  
CC at risk of getting cancer. Furthermore, the TADG-15 gene can be used as a  
CC diagnostic or therapeutic target in cancer. The present sequence was used  
CC in a sequence homology alignment with the catalytic domain of TADG-15.

XX Sequence 231 AA;  
SQ

Query Match 82.0%; Score 82; DB 22; Length 231;  
Best Local Similarity 78.9%; Pred. No. 8.1e-06;  
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEAVPHXWQVSLQ 19  
Db 2 IVNGEDAVPGSPWQVSLQ 20

RESULT 12  
AAU82738  
ID AAU82738 standard; Protein; 263 AA.  
XX  
AC AAU82738;  
XX  
XX 23-APR-2002 (first entry)  
XX Amino acid sequence of novel human protease #37.  
DE  
XX Human; protease; cancer; immune-related disorder; cardiovascular disease;  
KW neuronal-associated disease; metabolic disorder; inflammatory disorder;  
KW nervous system disorder; sexual dysfunction; pain; mood disorder;  
KW hypertension; psychotic disorder; neurological disorder; dyskinesia;  
KW viral infection; human immunodeficiency virus; HIV; non-viral infection;  
KW ocular disease; cytostatic; enzyme.  
XX  
OS Homo sapiens.  
XX WO200200860-A2.  
PN  
XX 03-JAN-2002.  
PD  
XX 26-JUN-2001; 2001WO-US20171.  
PF  
XX 26-JUN-2000; 2000US-214047P.  
PR  
XX (SUGS-) SUGEN INC.  
PA  
XX Plowman G, Whyte D, Sudarsanam S, Manning G, Caenepeel S;  
PI Charyczak G;  
PI  
XX WPI; 2002-139913/18.  
DR  
XX N-PSDS; ABK31780.  
DR  
XX Nucleic acids encoding novel human proteases, useful for useful for  
PT treating diseases and disorders such as cancers, immune-related  
PT diseases and disorders, cardiovascular disease (e.g. restenosis) and  
PT inflammatory disorders -  
PT  
XX  
PS Claim 6; Fig 2M; 313pp; English.  
XX  
CC The present invention relates to the isolation of novel human  
CC proteases, and the nucleic acids encoding them. The sequences of  
CC the invention are useful for treating diseases and disorders such as  
CC cancers (e.g. breast, colon, lung), immune-related diseases and disorders  
CC (e.g. inflammatory diseases and asthma), cardiovascular diseases  
CC (e.g. restenosis and coronary thrombosis), brain or neuronal-associated  
CC diseases, metabolic disorders (e.g. diabetes, obesity), inflammatory  
CC disorders (e.g. rheumatoid arthritis and psoriasis), central or  
CC peripheral nervous system diseases, migraines, pain, sexual dysfunction,  
CC mood disorders, attention disorders, neurological disorders  
CC hypertension, psychotic disorders, Parkinson's disease) and dyskinesias.  
CC The nucleic acids and polypeptides are also useful for treating viral  
CC infections caused by human immunodeficiency virus (HIV), and non-viral  
CC infections such as ocular disease (e.g. glaucoma) and macular  
CC degeneration. AAU82702-AAU82760 represent the novel human proteases of  
CC the invention.  
XX  
XX Sequence 263 AA;

```

Query Match      82.0%; Score 82; DB 23; Length 263;
Best Local Similarity 78.9%; Pred. No. 9.4e-06;
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSLQ 19
Db 34 IVNGEDAVPGSWPQVSLQ 52

RESULT 13
AAP60060
ID AAP60060 standard; Protein; 242 AA.
XX
AC AAP60060;
XX
DT 06-SEP-1991 (first entry)
XX
DE Sequence of human pancreatic elastase IIIA.
KW Enzyme; serum lipoprotein metabolism.
XX
OS Homo sapiens.
XX
PN EP198645-A.
XX
PD 22-OCT-1986.
XX
PF 07-APR-1986; 86EP-0302557.
XX
PR 23-OCT-1985; 85JP-0236586.
XX
PR 05-APR-1985; 85JP-0072308.
XX
PR 27-APR-1985; 85JP-0091986.
XX
PR 26-JUL-1985; 85JP-0163964.
XX
PR 02-DEC-1985; 85JP-0271128.
XX
PA (SANY ) SANKYO KK.
XX
PI Takiguchi Y, Tani T, Kawashima I, Erukawa H, Ohmine T;
PI Ohsumi J;
XX
DR WFI; 1986-280300/43.
DR N-PSDB; AAN60074.
XX
PT New genetically engineered human pancreatic elastase - obtd.
XX
PS Claim 2; Page 6-7; 45pp; English.
XX
CC The genetically engineered prod. can eliminate the dependency on
CC human pancreas supplies for the elastase, and avoids antibody
CC formation and possibility of anaphylaxis using porcine elastase.
XX
SQ Sequence 242 AA;

Query Match      80.0%; Score 80; DB 7; Length 242;
Best Local Similarity 68.4%; Pred. No. 1.9e-05;
Matches 13; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSLQ 19
Db 1 VHGEDAVPGSWPQVSLQ 19

RESULT 14
AAB11711
ID AAB11711 standard; Protein; 264 AA.
XX
AC AAB11711;
XX
DT 23-OCT-2000 (first entry)
XX
DE Mouse serine protease BSSP5 (mBSSP5) SEQ ID NO:4.

BSSP5; serine protease; human; hBSSP5; mouse; mBSSP5; brain;
diagnostic marker; antibody; transgenic animal; Alzheimer's disease;
epilepsy; cancer; inflammation; infertility; pancreatitis;
prostatic hypertrophy.
Mus sp.
WO2000031243-A1.
XX
EN 02-JUN-2000.
XX
PF 19-NOV-1999; 99WO-JP06473.
XX
PR 20-NOV-1998; 98JP-0347806.
XX
PA (FUSO ) FUSO PHARM IND LTD.
XX
PI Uemura H, Okui A, Kominami K, Yamaguchi N, Mitsui S;
XX
DR WPI; 2000-400058/34.
DR N-PSDB; AAA61734.
XX
PT Serine proteases BSSP5, useful in detecting homologs, mutants and
PT polymorphic variants as markers for diagnosis of e.g. Alzheimer's
PT disease, epilepsy, cancer and inflammation, using blood, urine,
PT pancreas or other tissues.
XX
PS Claim 3; Page 55-56; 70pp; Japanese.
XX
CC The invention relates to novel serine proteases designated BSSP5
CC (AAB11710-B11711), and to nucleic acids encoding them (AAA61733-A61734).
CC The invention also relates to vectors and transformants comprising BSSP5
CC nucleic acids; transgenic animals in which the expression level of BSSP5
CC can be varied; and an mBSSP5 knockout mouse. The invention additionally
CC encompasses anti-BSSP5 antibodies and methods of production of such
CC antibodies, methods of BSSP5 detection using the antibodies, and the
CC use of BSSP5 proteins or fragments as diagnostic markers for certain
CC medical conditions, e.g., pancreatitis. A method for detecting
CC pancreatitis comprising measuring BSSP5 concentration in the blood or
CC urine, and a pancreatitis diagnostic agent containing an anti-BSSP5
CC antibody is also disclosed. Nucleotides encoding BSSP5 were initially
CC isolated in a human brain cDNA library using degenerate PCR primers
CC (AAA61744-A61745) based on conserved regions of serine proteases. The
CC BSSP5 serine proteases and nucleotides encoding them are useful in
CC detecting homologues, mutants and polymorphic variants in biological
CC samples (e.g., blood, urine, brain, prostate gland, placenta, testis,
CC pancreas and spleen) as diagnostic markers for conditions such as
CC Alzheimer's disease, epilepsy, cancer, inflammation, infertility
CC pancreatitis and prostatic hypertrophy. Sequence AAB11710 represents
CC human BSSP5 (hBSSP5), and sequence AAB11711 represents murine BSSP5
CC (mBSSP5).
XX
SQ Sequence 264 AA;

Query Match      80.0%; Score 80; DB 21; Length 264;
Best Local Similarity 78.9%; Pred. No. 2.1e-05;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSLQ 19
Db 34 IVNGEDAVPGSWPQVSLQ 52

RESULT 15
AAP70759
ID AAP70759 standard; Protein; 270 AA.
XX
AC AAP70759;
XX
DT 19-APR-1991 (first entry)
XX
DE Human pancreas elastase-1.

```

Wed Feb 12 11:59:37 2003

```
XX Lipoprotein metabolism; lipid.
KW
XX
XX Sus scrofa.
XX
XX Key Location/Qualifiers
FH
FT Misc-difference 2..28
FT /note= "May be absent"
XX
XX JP62000276-A.
XX
XX 06-JAN-1987.
XX
XX 25-JUN-1985; 85JP-0138494.
XX
XX 25-JUN-1985; 85JP-0138494.
XX
XX (SANY ) SANKYO KK.
XX
XX WPI; 1987-040875/06.
XX N-PSDB; AAN71122.
XX
XX Pancreas elastase prodn. used to improve lipoprotein metabolism -
PT comprises isolating RNA coding elastase, synthesising single and
PT double chain cDNA and introducing recombinant into host
XX
XX PS Disclosure; Page 481; 18pp; Japanese.
XX
XX Sequence encoding pancreas elastase may be used to transform an
CC expression system. The product is useful in rejuvenating the
CC elasticity of the arterial wall, and improving unusual serum lipid
CC levels and lipoprotein metabolism.
XX
XX SQ Sequence 270 AA;
XX
XX Query Match 80.0%; Score 80; DB 8; Length 270;
XX Best Local Similarity 68.4%; Pred. No. 2.le-05;
XX Matches 13; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
Qy 1 IVNGEAVPHXWQVSLQ 19
Db 29 VVHGDAVPYSWPQVSLQ 47
Search completed: February 12, 2003, 10:22:24
Job time : 14.4328 secs
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:17:50 ; Search time 5.01493 Seconds  
(without alignments)  
383.393 Million cell updates/sec

Title: US-10-036-371-4

Perfect score: 100

Sequence: 1 IVNGEEAVPHXWQVSLQ 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 73:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97	97.0	263	S47537	chymotrypsin (EC 3
2	95	95.0	244	S72219	chymotrypsin B - A
3	85	85.0	31	B3257	pancreatic elastas
4	85	85.0	245	KYBOA	chymotrypsin (EC 3
5	85	85.0	270	B29934	pancreatic elastas
6	83	83.0	23	P00036	serine proteinase
7	83	83.0	28	A61529	chymotrypsin (EC 3
8	82	82.0	245	KYBOB	chymotrypsin (EC 3
9	82	82.0	263	A21195	chymotrypsin (EC 3
10	82	82.0	263	A31299	chymotrypsin (EC 3
11	81	81.0	263	KYRTB	chymotrypsin (EC 3
12	80	80.0	20	B61333	chymotrypsin (EC 3
13	80	80.0	268	JQ1473	pancreatic elastas
14	80	80.0	270	A29934	pancreatic elastas
15	77	77.0	268	S68826	pancreatic elastas
16	77	77.0	268	S68825	pancreatic elastas
17	75	75.0	240	CPBOA3	pancreatic elastas
18	74	74.0	269	B26823	procarboxypeptidas
19	74	74.0	271	A25528	pancreatic elastas
20	73	73.0	126	A23473	pancreatic elastas
21	72	72.0	271	ELRT2	chymotrypsin-like
22	71	71.0	269	A26823	pancreatic elastas
23	70	70.0	264	I38136	pancreatic elastas
24	69	69.0	269	C26823	chymotrypsin-like
25	67	67.0	1004	T30338	oviductin (EC 3.4
26	65	65.0	20	S33787	pancreatic elastas
27	65	65.0	25	S19997	pancreatic elastas
28	65	65.0	1524	T30337	polyprotein - Afri
29	64	64.0	460	B61545	plasmin (EC 3.4.21

#### ALIGNMENTS

##### RESULT 1

S47537  
chymotrypsin (EC 3.4.21.1) precursor - Atlantic cod  
C:Species: Gadus morhua (Atlantic cod)  
C:Date: 26-Dec-1994 #sequence\_revision 03-Aug-1995 #text\_change 22-Jun-1999  
C:Accession: S47537; S43163  
R:Gudmundsdottir, A.; Oskarsson, S.; Eakin, A.E.; Craik, C.S.; Bjarnason, J.B.  
Biochim. Biophys. Acta 1219, 211-214, 1994  
A:Title: Atlantic cod cDNA encoding a psychrophilic chymotrypsinogen.  
A:Reference number: S47537, MUID:94368860; PMID:8086467  
A:Accession: S47537  
A:Molecule type: mRNA  
A:Residues: 1-263 <GUD>  
A:Cross-references: EMBL:X78490; NID:g468750; PIDN:CAA55242.1; PID:g468751  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; protein digestion; serine proteinase  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-263/Product: chymotrypsin #status predicted <SIG>  
F:34-256/Domain: trypsin homology <TRY>  
F:75,120,213/Active site: His, Asp, Ser #status predicted

Query Match 97.0%; Score 97; DB 2; Length 263;  
Best Local Similarity 89.5%; Pred. No. 1.2e-08;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1	IVNGEEAVPHXWQVSLQ	19
Db	34	IVNGEEAVPHXWQVSLQ	52

##### RESULT 2

S72219  
chymotrypsin B - Atlantic cod (fragments)  
C:Species: Gadus morhua (Atlantic cod)  
C:Date: 19-Mar-1998 #sequence\_revision 17-Apr-1998 #text\_change 07-Aug-1998  
C:Accession: S72219  
R:Leth-Larsen, R.; Asgeirsson, B.; Thorolfsson, M.; Norregaard-Madsen, M.; Hojrup, P.  
Biochim. Biophys. Acta 1297, 49-56, 1996  
A:Title: Structure of chymotrypsin variant B from Atlantic cod, Gadus morhua.  
A:Reference number: S72219; MUID:96439045; PMID:8841380  
A:Accession: S72219  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-14,15-244 <LET>  
C:Superfamily: trypsin; trypsin homology  
F:15-237/Domain: trypsin homology <TRY>

Query Match 95.0%; Score 95; DB 2; Length 244;  
Best Local Similarity 89.5%; Pred. No. 2.4e-08;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1	IVNGEEAVPHXWQVSLQ	19
----	---	-------------------	----

Db 15 IVNGEEAVPHSWPQVSLQ 33

|||||

RESULT 3

B33257

pancreatic elastase (EC 3.4.21.36) III - pig (fragment)

N;Alternate names: proteinase E

C;Species: Sus scrofa domestica (domestic pig)

C;Date: 21-Feb-1990 #sequence\_revision 21-Feb-1990 #text\_change 30-Jun-1995

C;Accession: B33257

R;Aviles, F.X.; Pascual, R.; Salva, M.; Bonicel, J.; Puigserver, A.

Biochem. Biophys. Res. Commun. 163, 1191-1196, 1989

A;Title: Generation of a subunit III-like protein by autolysis of human and porcine pro-

A;Reference number: A33257; PMID:89392022; PMID:2675835

A;Accession: B33257

A;Status: preliminary

A;Molecule type: protein

A;Residues: 1-31 <AVI>

C;Superfamily: trypsin; trypsin homology

C;Keywords: hydrolase; serine proteinase

Query Match 85.0%; Score 85; DB 2; Length 31;

Best Local Similarity 73.7%; Pred. No. 1.2e-07;

Matches 14; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHSWPQVSLQ 19

|||||

Db 12 VVNGEDAVPSWNPQVSLQ 30

|||||

RESULT 4

KYBOA

chymotrypsin (EC 3.4.21.1) A precursor - bovine

N;Alternate names: chymotrypsinogen A

C;Species: Bos primigenius taurus (cattle)

C;Date: 07-May-1981 #sequence\_revision 07-May-1981 #text\_change 07-May-1999

C;Accession: A90235; A93158; S29650; A00952

R;Brown, J.K.; Hartley, B.S.

Biochem. J. 101, 214-228, 1966

A;Title: Location of disulphide bridges by diagonal paper electrophoresis. The disulphid

A;Reference number: A90235; PMID:67181721; PMID:5971783

A;Accession: A90235

A;Molecule type: protein

A;Residues: 1-101, N', 103-245 <BRO>

R;Blow, D.M.; Birktoft, J.J.; Hartley, B.S.

Nature 221, 337-340, 1969

A;Title: Role of a buried acid group in the mechanism of action of chymotrypsin.

A;Reference number: A93158; PMID:69106266; PMID:5764436

A;Contents: annotation; revision to residue 102

R;Meloun, B.; Klueh, I.; Kostka, V.; Moravek, L.; Prusik, Z.; Vanacek, J.; Keil, B.; Sorn

Biochim. Biophys. Acta 130, 543-546, 1966

A;Title: Covalent structure of bovine chymotrypsinogen A.

A;Reference number: A90572; PMID:67183948; PMID:5972866

A;Accession: A93158

A;Molecule type: protein

A;Residues: 1-101, N', 103-245 <MEL>

A;Note: disulfide bonds were determined

R;Cruzolola, F.; Ascenzi, P.; Barra, D.; Bolognesi, M.; Menegatti, E.; Sarti, P.; Schne

Biochim. Biophys. Acta 1161, 201-209, 1993

A;Title: Selective oxidation of Met-192 in bovine alpha-chymotrypsin. Effect on catalyti

A;Reference number: S29650; PMID:93160238; PMID:8431470

A;Accession: S29650

A;Molecule type: protein

A;Residues: 1-12; 16-27; 149-160; 181-200 <CUT>

R;Smillie, L.B.; Hartley, B.S.

Biochem. J. 101, 232-241, 1966

A;Title: Histidine sequences in the active centres of some 'serine' proteinases.

A;Reference number: A90236; PMID:67181723; PMID:5971785

A;Contents: annotation; active site

R;Birktoft, J.J.; Blow, D.M.; Henderson, R.; Steitz, T.A.

Philos. Trans. R. Soc. Lond. B257, 67-76, 1970

A;Title: The structure of alpha-chymotrypsin.

A;Reference number: A93754

A;Contents: annotation; X-ray crystallography

C;Comment: Chymotrypsinogens are synthesized in the acinar cells of pancreas.

C;Comment: Tryptic cleavage after Arg-15 results in a fully active enzyme (pi-chymotrypsin)

d Asn-148 directly from chymotrypsinogen, which leads to the degraded form neochymotrypsin

C;Superfamily: trypsin; trypsin homology

C;Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen

F;1-245/Product: chymotrypsinogen #status experimental <ZYM>

F;1-238/Domains: trypsin homology <TRY>

F;1-122, 42-58, 136-201, 168-182, 191-220/Disulfide bonds: #status experimental

F;57, 102, 195/Active site: His, Asp, Ser #status experimental

Query Match 85.0%; Score 85; DB 1; Length 245;

Best Local Similarity 84.2%; Pred. No. 1.1e-06;

Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHSWPQVSLQ 19

|||||

Db 16 IVNGEEAVPGSWNPQVSLQ 34

|||||

RESULT 5

B29934

pancreatic elastase (EC 3.4.21.36) IIB precursor - human

N;Alternate names: 35K glycoprotein; pancreatic protein P35; protein G32; proteinase E

C;Species: Homo sapiens (man)

C;Date: 15-Dec-1988 #sequence\_revision 15-Dec-1988 #text\_change 22-Jun-1999

C;Accession: B29934; A90516; A27206; A33257; A28932; S04399; S04490

R;Tani, T.; Ohsumi, J.; Mita, K.; Takiguchi, Y.

J. Biol. Chem. 263, 12311-1239, 1988

A;Title: Identification of a novel class of elastase isozyme, human pancreatic elastase J

A;Reference number: A92664; PMID:88087253; PMID:2826474

A;Accession: B29934

A;Molecule type: mRNA

A;Residues: 1-270 <TAN>

A;Cross-references: GB:M18692; NID:9607029; PIDN:AAA58454.1; PID:gl82035

R;Shen, W.; Fletcher, T.S.; Largman, C.

Biochemistry 26, 3447-3452, 1987

A;Title: Primary structure of human pancreatic protease E determined by sequence analysis

A;Reference number: A90516; PMID:88000545; PMID:3477287

A;Accession: A90516

A;Molecule type: mRNA

A;Residues: 'G', 5-63, 'G', 65-78, 'W', 80-118, 'G', 120-163, 'P', 165-270 <SHE>

R;Fletcher, T.S.

submitted to GenBank, August 1987

A;Reference number: A94507

A;Contents: revision to residue 119

A;Accession: A27206

A;Molecule type: DNA

A;Residues: 'G', 5-63, 'G', 65-78, 'W', 80-163, 'P', 165-270 <FLE>

R;Aviles, F.X.; Pascual, R.; Salva, M.; Bonicel, J.; Puigserver, A.

Biochem. Biophys. Res. Commun. 163, 1191-1196, 1989

A;Title: Generation of a subunit III-like protein by autolysis of human and porcine pro

A;Reference number: A33257; PMID:89392022; PMID:2675835

A;Accession: A33257

A;Molecule type: protein

A;Residues: 18-57 <AVI>

R;Guy-Crotte, O.; Barthe, C.; Basso, D.; Fournet, B.; Figarella, C.

Biochem. Biophys. Res. Commun. 156, 318-322, 1988

A;Title: Characterization of two glycoproteins of human pancreatic juice: P35, a truncate

A;Reference number: A28932; PMID:89025862; PMID:3178837

A;Accession: A28932

A;Molecule type: protein

A;Residues: 'X', 32-52, 'X', 54-55, 'XXX', 59-63 <GUY>

R;Moulard, M.; Kerfelec, B.; Mallet, B.; Chapuis, C.

FEBS Lett. 250, 166-170, 1989

A;Title: Identification of a procaboxypeptidase A-truncated protease E binary complex ir

A;Reference number: S04999; PMID:89325560; PMID:2753124

A;Accession: S04999

A;Molecule type: protein

A;Residues: 31-50 <MOU>

R.Wendorf, P.; Geyer, R.; Sziegoleit, A.; Linder, D.  
 FEBS Lett. 249, 275-278, 1989

A:Title: Localization and characterization of the glycosylation site of human pancreatic  
 A:Reference number: S04490; MUID:89289996; PMID:2737288

A:Accession: S04490

A:Molecule type: protein

A:Residues: 94-128:132-164 <WEN>

C:Superfamily: trypsin; trypsin homology

C:Keywords: glycoprotein; hydrolase; serine proteinase; zymogen

F:1-17/Domain: signal sequence #status predicted <SIG>

F:18-28/Domain: activation peptide #status predicted <ACT>

F:29-270/Product: (or 31-270) pancreatic elastase IIIB #status predicted <MAT>

F:29-263/Domain: trypsin homology <TRY>

F:73-123,217/Active site: His, Asp, Ser #status predicted

F:114/Binding site: carbohydrate (Asn) (covalent) #status experimental

F:153/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 85.0%; Score 85; DB 2; Length 270;

Best Local Similarity 73.7%; Pred. No. 1.2e-06;

Matches 14; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVNGEAVPHXKXWQVSLQ 19

DB 29 VVNGDAVPSWQVSLQ 47

RESULT 6

PU0036

N:Alternative names: trypsin-like proteinase bPTIP

C:Species: Bos primigenius taurus (cattle)

C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 22-Apr-1995

C:Accession: PU0036; PU0039

R:Tsujii, A.; Sakiyama, K.; Edazawa, K.; Nagata, K.; Sasaki, Y.; Nagamune, H.; Matsuda, Y.

submitted to JIPID, September 1994

A:Description: Purification and characterization of a novel serine proteinase from bovin

A:Reference number: PU0036

A:Accession: PU0036

A:Molecule type: protein

A:Residues: 1-23 <TSU>

A:Experimental source: pancreas

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; serine proteinase

Query Match 83.0%; Score 83; DB 2; Length 23;

Best Local Similarity 63.2%; Pred. No. 1.8e-07;

Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEAVPHXKXWQVSLQ 19

DB 1 VVGEDAIPHSPWQVSLQ 19

RESULT 7

A61529

N:Alternative names: Atlantic cod (fragments)

C:Species: Gadus morhua (Atlantic cod)

C:Date: 15-Oct-1994 #sequence\_revision 15-Oct-1994 #text\_change 11-May-2000

C:Accession: A61529

R:Aggeirsson, B.; Bjarnason, J.B.

Comp. Biochem. Physiol. B 99, 327-335, 1991

A:Title: Structural and kinetic properties of chymotrypsin from Atlantic cod (Gadus morh

A:Reference number: A61529; MUID:92111252; PMID:1764912

A:Accession: A61529

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-28 <ASG>

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; protein digestion; serine proteinase

Query Match

Best Local Similarity 83.0%; Score 83; DB 2; Length 28;

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVNGEAVPHXKXWQV 16

DB 13 IVNGEAVPHRTWYQV 28

RESULT 8

KYBOB

N:Alternative names: chymotrypsinogen B

C:Species: Bos primigenius taurus (cattle)

C:Date: 08-Oct-1981 #sequence\_revision 08-Oct-1981 #text\_change 18-Jul-1997

C:Accession: A00953

R:Smillie, L.B.; Furka, A.; Nagabhushan, N.; Stevenson, K.J.; Parkes, C.O.

Nature 218, 343-346, 1968

A:Title: Structure of chymotrypsinogen B compared with chymotrypsinogen A and trypsinogen

A:Reference number: A00953; MUID:68238908; PMID:5649671

A:Accession: A00953

A:Molecule type: protein

A:Residues: 1-245 <SMI>

C:Comment: Chymotrypsinogen B is synthesized, along with chymotrypsinogen A, in the acine

C:Comment: The first activation cleavage, leading to pi-chymotrypsin B, occurs in the ear

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen

F:1-15/Domain: propeptide #status experimental <PRO>

F:16-245/Product: chymotrypsin B #status experimental <MAT>

F:16-238/Domain: trypsin homology <TRY>

F:1-122,42-58,136-201,168-182,191-220/Dissulfide bonds: #status experimental

F:57,102,195/Active site: His, Asp, Ser #status experimental

Query Match

Best Local Similarity 82.0%; Score 82; DB 1; Length 245;

Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEAVPHXKXWQVSLQ 19

DB 16 IVNGDAVPGSWQVSLQ 34

RESULT 9

A21195

N:Alternative names: 2 precursor - dog

C:Species: Canis lupus familiaris (dog)

C:Date: 27-Jul-1990 #sequence\_revision 27-Jul-1990 #text\_change 22-Jun-1999

C:Accession: A21195

R:Pinsky, S.D.; LaForge, K.S.; Luc, V.; Scheele, G.

Proc. Natl. Acad. Sci. U.S.A. 80, 7486-7490, 1983

A:Title: Identification of cDNA clones encoding secretory isoenzyme forms: sequence deter

A:Reference number: A21195; MUID:84170253; PMID:6584866

A:Accession: A21195

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-263 <PIN>

A:Cross-references: GB:K01173; NID:g163945; PIDN:AAA30841.1; PID:g163946

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; protein digestion; serine proteinase

F:34-256/Domain: trypsin homology <TRY>

F:75,120,213/Active site: His, Asp, Ser #status predicted

Query Match

Best Local Similarity 82.0%; Score 82; DB 2; Length 263;

Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEAVPHXKXWQVSLQ 19

DB 34 IVNGDAVPGSWQVSLQ 52

RESULT 10

A31299

N:Alternative names: 1 precursor - human

C:Species: Homo sapiens (man)

C:Date: 08-Jun-1989 #sequence\_revision 08-Jun-1989 #text\_change 22-Jun-1999





F:1-15/Domain: signal sequence #status predicted <SIG>  
F:16-28/Domain: propeptide #status predicted <PRO>  
F:29-270/Product: pancreatic elastase II/IIA #status experimental <MAT>  
F:29-263/Domain: trypsin homology <TRY>  
F:73,123,217/Active site: His, Asp Ser #status predicted

Query Match 80.0%; Score 80; DB 2; Length 270;  
Best Local Similarity 68.4%; pred. No. 8.6e-06;  
Matches 13; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

**Qy** 1 IVNGEEAVPHXWXWQVSLQ 19  
:  
**Dd** 29 VVHGDAVPYSPWPQVSLQ 47

RESULT 15

S68826  
pancreatic elastase (EC 3.4.21.36) isoform 2 precursor - human  
N/Alternate names: caldesin isoform 2  
C/Species: Homo sapiens (man)  
C/Date: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #text\_change 30-Jan-1998  
C/Accession: S68826  
R/Tomomura, A.; Akiyama, M.; Itoh, H.; Yoshino, I.; Tomomura, M.; Nishii, Y.; Noikura, T.  
FEBS Lett. 386, 26-28, 1996  
A/Title: Molecular cloning and expression of human caldesin.  
A/Reference number: S68825; MUID:96221265; PMID:8635596  
A/Accession: S68826  
A/Molecule type: mRNA  
A/Residues: 1-268 <TOM>  
A/Experimental source: pancreas  
C/Superfamily: trypsin; trypsin homology  
C/Keywords: hydrolase; pancreas; serine proteinase; zymogen  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-29/Domain: propeptide #status predicted <PRO>  
F:30-268/Product: pancreatic elastase isoform 2 #status predicted <MAT>  
F:30-262/Domain: trypsin homology <TRY>

Query Match 77.0%; Score 77; DB 2; Length 268;  
Best Local Similarity 63.2%; Pred. No. 2.7e-05;  
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY            1 IVNGEAVPHXWXWQVSLQ 19  
               :|:|||||:||||  
Db            30 VVGGEDARPHSWPWQISLQ 48

Search completed: February 12, 2003, 10:28:49  
Job time : 5.01493 secs

**THIS PAGE BLANK (USP 11)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:04:45 ; Search time 2.68657 Seconds  
(without alignments)  
308.768 Million cell updates/sec

Title: US-10-036-371-4

Perfect score: 100

Sequence: 1 IVNGEAVPHXWQVSLQX 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97	97.0	263	1 CTRA_GADMO	P47796 gadus morhu
2	95	95.0	245	1 CTRB_GADMO	P80646 gadus morhu
3	87	87.0	253	1 CAC3_BOVIN	P05805 bos taurus
4	85	85.0	245	1 CTRA_BOVIN	P00766 bos taurus
5	85	85.0	245	1 EL3B_BOVIN	P08861 homo sapien
6	82	82.0	245	1 CTRB_BOVIN	P00767 bos taurus
7	82	82.0	263	1 CTR2_CANFA	P04813 canis fami
8	82	82.0	263	1 CTRB_HUMAN	P17538 homo sapien
9	81	81.0	263	1 CTRB_RAT	P07338 homo sapien
10	80	80.0	270	1 EL3A_HUMAN	P09093 homo sapien
11	78	78.0	268	1 CLCR_RAT	P55091 rattus norv
12	77	77.0	268	1 CLCR_HUMAN	P09895 homo sapien
13	74	74.0	269	1 EL2A_HUMAN	P08217 homo sapien
14	74	74.0	271	1 EL2_MOUSE	P05208 mus musculu
15	72	72.0	271	1 EL2_RAT	P00774 rattus norv
16	71	71.0	269	1 EL2_BOVIN	P29461 bos taurus
17	71	71.0	269	1 EL2_PIG	P08419 sus scrofa
18	70	70.0	264	1 CTR1_HUMAN	P40313 homo sapien
19	69	69.0	269	1 EL2B_HUMAN	P08218 homo sapien
20	66	66.0	492	1 TMS2_HUMAN	O15393 homo sapien
21	65	65.0	20	1 ELAS_GADMO	P13297 gadus morhu
22	65	65.0	275	1 TRYPT_PIG	Q9n2d1 sus scrofa
23	64	64.0	343	1 PLMN_SHEEP	P81286 ovis aries
24	63	63.0	790	1 PLMN_PIG	P06867 sus scrofa
25	62	62.0	20	1 COG1_CHTOP	P34153 chionocete
26	62	62.0	437	1 TMS4_HUMAN	Q9nrs4 homo sapien
27	62	62.0	810	1 PLMN_MACMU	P12545 macaca mula
28	62	62.0	1420	1 APOA_MACMU	P14417 macaca mula
29	62	62.0	4548	1 APOA_HUMAN	P08519 homo sapien
30	61	61.0	810	1 PLMN_HUMAN	P00747 homo sapien
31	60	60.0	333	1 PLMN_CANFA	P80009 canis fami
32	60	60.0	812	1 PLMN_BOVIN	P06868 bos taurus
33	60	60.0	812	1 PLMN_MOUSE	P20918 mus musculu

34	59	59.0	270	1 TRYT_MERUN	P50342 meriones un
35	59	59.0	273	1 MCT7_RAT	P27435 rattus norv
36	59	59.0	338	1 PLMN_HORSE	P80010 equus cabal
37	58	58.0	273	1 MCT7_MOUSE	Q02844 mus musculu
38	58	58.0	810	1 PLMN_ERIEU	Q29485 erinaceus e
39	57	57.0	235	1 TRYD_HUMAN	Q9B2J3 homo sapien
40	57	57.0	267	1 ELNE_HUMAN	P08246 homo sapien
41	57	57.0	275	1 TRB1_HUMAN	Q15661 homo sapien
42	57	57.0	275	1 TRB2_HUMAN	P20231 homo sapien
43	57	57.0	275	1 TRYA_HUMAN	P15157 homo sapien
44	56	56.0	254	1 KLK4_HUMAN	Q9Y5K2 homo sapien
45	56	56.0	274	1 MCT6_RAT	P50343 rattus norv

#### ALIGNMENTS

```

RESULT 1
CTRA_GADMO
ID CTRA_GADMO STANDARD; PRT; 263 AA.
AC P47796;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymotrypsin A precursor (RC 3.4.21.1).
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.
OX NCBI_TaxID=8049;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pyloxic caeca;
RX MEDLINE=9436860; PubMed=8086467;
RA Gudmundsdottir A., Oskarsson S., Bakin A.E., Craik C.S.,
RA Bjarnason J.B.
RT "Structural and kinetic properties of chymotrypsin from Atlantic cod
  (Gadus morhua). Comparison with bovine chymotrypsin.";
RL Comp. Biochem. Physiol. 99B:327-335(1991).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr|-Xaa, Trp|-Xaa,
  Phe|-Xaa, Leu|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to licenses@isb-sib.ch).
-----
EMBL; X78490; CAA55242.1; -.
DR HSSP; P00766; 1CHG.
DR MEROPS; S01.152; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal.
FT SIGNAL 1 18
FT CHAIN 19 263 CHYMOTRYPSIN A.

```

```
FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 19 140 BY SIMILARITY.
FT DISULFID 60 76 BY SIMILARITY.
FT DISULFID 154 219 BY SIMILARITY.
FT DISULFID 186 200 BY SIMILARITY.
FT DISULFID 209 238 BY SIMILARITY.
FT DISULFID 21 21 R -> S (IN REF. 2).
FT CONFLICT 25 25 S -> Q (IN REF. 2).
FT CONFLICT 29 29 T -> S (IN REF. 2).
FT CONFLICT 44 44 S -> T (IN REF. 2).
FT CONFLICT 46 46 S -> Y (IN REF. 2).
SQ SEQUENCE 263 AA; 28294 MW; 47AAC699A0A64FBB CRC64;

Query Match 97.0%; Score 97; DB 1; Length 263;
Best Local Similarity 89.5%; Pred. No. 2.7e-08;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXKXWQVSLQ 19
Db 34 IVNGEEAVPHXKXWQVSLQ 52

RESULT 2
CTRB GADMO STANDARD; PRT; 245 AA.
AC P80646;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymotrypsin B (EC 3.4.21.1).
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.
OX NCBI_TaxID=8049;
RN [1]
RP TISSUE=Pyloric caeca;
RC MEDLINE=96439045; PubMed=8841380;
RA Leth-Larsen R., Asgeirsson B., Thorolfsson M., Noerregaard-Madsen M.,
RA Hostrup P.;
RT "Structure of chymotrypsin variant B from Atlantic cod, Gadus
RT morhua.";
RL Biochim. Biophys. Acta 1297:49-56(1996).
RN [2]
RP SEQUENCE OF 1-12 AND 16-31.
RC TISSUE=Pyloric caeca;
RX MEDLINE=92111252; PubMed=1764912;
RA Asgeirsson B., Bjarnason J.B.;
RT "Structural and kinetic properties of chymotrypsin from Atlantic cod
RT (Gadus morhua). Comparison with bovine chymotrypsin.";
RL Comp. Biochem. Physiol. 99B:327-335(1991).
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
CC Phe-|-Xaa, Leu-|-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC HSP; P00766; 1CHG.
DR MEROPS; S01.152; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser. protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; P00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen.
FT CHAIN 1 13 CHYMOTRYPSIN B, A CHAIN.
FT CHAIN 16 245 CHYMOTRYPSIN B, B CHAIN.
FT ACT_SITE 57 57 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 101 101 CHARGE RELAY SYSTEM (BY SIMILARITY).
```

```
FT ACT_SITE 195 195 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 1 121 BY SIMILARITY.
FT DISULFID 42 58 BY SIMILARITY.
FT DISULFID 135 201 BY SIMILARITY.
FT DISULFID 167 182 BY SIMILARITY.
FT DISULFID 191 220 BY SIMILARITY.
FT CONFLICT 9 11 QVT -> VIS (IN REF. 2).
FT CONFLICT 26 26 S -> T (IN REF. 2).
FT CONFLICT 28 29 PW -> Y (IN REF. 2).
SQ SEQUENCE 245 AA; 26260 MW; 74FE0D425517AB02 CRC64;

Query Match 95.0%; Score 95; DB 1; Length 245;
Best Local Similarity 89.5%; Pred. No. 5.2e-08;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXKXWQVSLQ 19
Db 16 IVNGEEAVPHXKXWQVSLQ 34

RESULT 3
CAC3_BOVIN STANDARD; PRT; 253 AA.
AC P05805;
DT 01-NOV-1988 (Rel. 09, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proproteinase B precursor (Procarboxypeptidase A complex component
DE III) (Procarboxypeptidase A-S6 subunit III) (PROCPA-S6 III).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE OF 1-25.
RX MEDLINE=91099520; PubMed=2269366;
RA Pascual R., Vendrell J., Aviles F.X., Bonicel J., Wicker C.,
RA Puigserver A.;
RT "Autolysis of proproteinase B in bovine procarboxypeptidase A ternary
RT complex gives rise to subunit III.";
RL FEBS Lett. 277:137-41(1990).
RN [2]
RP SEQUENCE OF 14-253, AND DISULFIDE BONDS.
RX MEDLINE=86220198; PubMed=3519215;
RA Venot N., Sciaky M., Puigserver A., Desnuelle P., Laurent G.;
RT "Amino acid sequence and disulfide bridges of subunit III, a
RT defective endopeptidase present in the bovine pancreatic 6 S
RT procarboxypeptidase A complex.";
RL Bur. J. Biochem. 157:91-99(1986).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RX MEDLINE=94222022; PubMed=8168476;
RA Pignol D., Gaboriaud C., Michon T., Kerfelec B., Chapus C.,
RA Fontecilla-Camps J.C.;
RT "Crystal structure of bovine procarboxypeptidase A-S6 subunit III, a
RT highly structured truncated zymogen E.";
RL EMBO J. 13:1763-1771(1994).
CC -1- FUNCTION: DEFECTIVE ELASTASE-LIKE SERINE PROTEASE. DOES NOT SEEM
CC TO HAVE A PROTEASE ACTIVITY. ITS LIKELY FUNCTION IS TO PROTECT
CC PROCARBOXYPEPTIDASE A AGAINST DENATURATION IN THE ACIDIC
CC ENVIRONMENT OF THE RUMINANT DUODENUM.
CC -1- SUBUNIT: HETEROTRIMER OF SUBUNIT III; CARBOXYPEPTIDASE A AND
CC CHYMOTRYPSINOGEN C.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR PIR; A25065; CPB0A3.
DR PDB; 1FON; 14-OCT-96.
DR MEROPS; S01.983; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser. protease_Try.
DR Pfam; PF00089; trypsin; 1.
```

```

DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SMO0020; TRYP_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Serine protease homolog; Pancreas; Digestion; 3D-structure.
FT PROPEP 1 11
FT CHAIN 12 253
FT DISULFID 41 57
FT DISULFID 100 103
FT DISULFID 140 206
FT DISULFID 171 187
FT DISULFID 196 227
SQ SEQUENCE 253 AA; 27337 MW; 24663724D8AE409C CRC64;

Query Match 87.0%; Score 87; DB 1; Length 253;
Best Local Similarity 73.7%; Pred. No. 9.9e-07;
Matches 14; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVNGEAVPHXWQVSLQ 19
Db 12 VVNGEDAVPYSWQVSLQ 30

RESULT 4
CTRA_BOVIN STANDARD; PRT; 245 AA.
AC P00766;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymotrypsinogen A (EC 3.4.21.1).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID:9913;
[1]
RN SEQUENCE, DISULFIDE BONDS, AND ACTIVE SITE.
RX MEDLINE=67181721; PubMed=5971783;
RA Brown J.R., Hartley B.S.;
RT "Location of disulphide bridges by diagonal paper electrophoresis.
RL Biochem. J. 101:214-228(1966).
RN [2]
RP REVISION TO 102.
RX MEDLINE=69106266; PubMed=5764436;
RA Blow D.M., Birktoft J.J., Hartley B.S.;
RT "Role of a buried acid group in the mechanism of action of
RL chymotrypsin.";
RN [3]
RP PRELIMINARY SEQUENCE.
RA Hartley B.S.;
RT "Amino-acid sequence of bovine chymotrypsinogen-A.";
RN [4]
RP SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE=67183948; PubMed=5972866;
RA Meloun B., Kluh I., Kostka V., Moravek L., Prusik Z., Vanacek J.,
RA Keil B., Sorm F.;
RT "Covalent structure of bovine chymotrypsinogen A.";
RL Biochim. Biophys. Acta 130:543-546(1966).
RN [5]
RP ACTIVE SITE.
RX MEDLINE=67181723; PubMed=5971785;
RA Smillie L.B., Hartley B.S.;
RT "Histidine sequences in the active centres of some 'serine'
RL proteinases.";
RN [6]
RP X-RAY CRYSTALLOGRAPHY.
RX MEDLINE=72035052; PubMed=4399050;

```

---

```

RA Birktoft J.J., Blow D.M., Henderson R., Steitz T.A.;
RT "1. Serine proteinases. The structure of alpha-chymotrypsin.";
RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 257:67-76(1970).
RN [7]
RX X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF CHYMOTRYPSINOGEN.
RA Freer S.T., Kraut J., Robertus J.D., Wright H.T., Xuong N.H.;
RT "Chymotrypsinogen: 2.5-A crystal structure, comparison with alpha-
chymotrypsin, and implications for zymogen activation.";
RN Biochemistry 9:1997-2009(1970).
RN [8]
RX X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF GAMMA-CHYMOTRYPSIN.
RA Cohen G.H., Silvertown E.W., Davies D.R.;
RT "Refined crystal structure of gamma-chymotrypsin at 1.9-A resolution.
Comparison with other pancreatic serine proteases.";
RN J. Mol. Biol. 148:449-479(1981).
RN [9]
RX X-RAY CRYSTALLOGRAPHY (1.68 ANGSTROMS) OF ALPHA-CHYMOTRYPSIN.
RA Tsukada H., Blow D.M.;
RT "Structure of alpha-chymotrypsin refined at 1.68-A resolution.";
RL J. Mol. Biol. 184:703-711(1985).
CC -|- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
CC Phe-|-Xaa, Leu-|-Xaa.
CC -|- SUBCELLULAR LOCATION: Extracellular.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -|- DATABASE: NAME=Worthington enzyme manual;
CC WWW="http://www.worthington-biochem.com/manual/C/CHY.html".
DR PIR; A00952; KYBOA.
DR PDB; 2CGA; 15-APR-90.
DR PDB; 2CHA; 31-MAY-84.
DR PDB; 4CHA; 29-OCT-85.
DR PDB; 5CHA; 16-OCT-87.
DR PDB; 6CHA; 16-OCT-87.
DR PDB; 1CHG; 27-JAN-84.
DR PDB; 1CHO; 16-JUL-88.
DR PDB; 2GCH; 31-MAY-84.
DR PDB; 3GCH; 15-OCT-92.
DR PDB; 4GCH; 15-OCT-90.
DR PDB; 5GCH; 15-OCT-90.
DR PDB; 6GCH; 15-OCT-90.
DR PDB; 7GCH; 15-OCT-90.
DR PDB; 8GCH; 15-JUL-93.
DR PDB; 1GCT; 15-OCT-91.
DR PDB; 2GCT; 15-OCT-91.
DR PDB; 3GCT; 15-OCT-91.
DR PDB; 1ACB; 31-OCT-93.
DR PDB; 1GMC; 31-OCT-93.
DR PDB; 1GMD; 31-OCT-93.
DR PDB; 1CGI; 30-APR-94.
DR PDB; 1CGJ; 30-APR-94.
DR PDB; 1GCD; 22-JUN-94.
DR PDB; 1GHA; 22-JUN-94.
DR PDB; 1GHB; 22-JUN-94.
DR PDB; 1GMH; 30-SEP-94.
DR PDB; 2GMT; 01-NOV-94.
DR PDB; 1MTN; 17-AUG-96.
DR PDB; 1AB9; 20-AUG-97.
DR PDB; 1AFO; 17-SEP-97.
DR PDB; 1CAO; 23-JUL-97.
DR PDB; 1CBW; 23-JUL-97.
DR PDB; 1VGC; 12-NOV-97.
DR PDB; 2VGC; 12-NOV-97.
DR PDB; 3VGC; 12-NOV-97.
DR PDB; 4VGC; 12-NOV-97.
DR PDB; 1HJA; 14-JAN-98.
DR MEROPS; S01.001; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser. protease_Try.
DR Pfam; PF00089; trypsin_1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SMO0020; TRYP_SPC; 1.

```



```

DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser. protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Zymogen; Signal; Glycoprotein.
FT SIGNAL 1 15 (POTENTIAL).
FT PROPEP 16 28 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 29 270 ELASTASE IIIB.
FT ACT_SITE 73 73 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 123 123 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 58 74 BY SIMILARITY.
FT DISULFID 117 120 PROBABLE.
FT DISULFID 157 223 BY SIMILARITY.
FT DISULFID 188 204 BY SIMILARITY.
FT DISULFID 213 244 BY SIMILARITY.
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .).
FT CONFLICT 4 4 R -> G (IN REF. 3).
FT CONFLICT 64 64 A -> G (IN REF. 3).
FT CONFLICT 79 79 W -> R (IN REF. 1).
FT CONFLICT 129 131 MISSING (IN REF. 5).
FT CONFLICT 164 164 R -> P (IN REF. 3).
SQ SEQUENCE 270 AA; 29293 MW; B14BE0AD3695AFE CRC64;

Query Match 85.0%; Score 85; DB 1; Length 270;
Best Local Similarity 73.7%; Pred. No. 2.2e-06;
Matches 14; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 IVNGEEAVPHXKXWQVSLQ 19
:|||||:|:|:|
DB 29 VVNGEDAVPGSWPQVSLQ 47

RESULT 6
CTRB_BOVIN
ID CTRB_BOVIN STANDARD; PRT; 245 AA.
AC P00767;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymotrypsinogen B (EC 3.4.21.1).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE, DISULFIDE BONDS, AND ACTIVE SITE.
RX MEDLINE=68238908; PubMed=5649671;
RA Smillie L.B., Furka A., Nagabhushan N., Stevenson K.J., Parkes C.O.;
RT "Structure of chymotrypsinogen B compared with chymotrypsinogen A and
trypsinogen.";
RL Nature 218:343-346(1968).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
Phe-|-Xaa, Leu-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- DATABASE: NAME=Worthington enzyme manual;
WWW="http://www.worthington-biochem.com/manual/C/CHY.html".
DR PIR; A00953; KYBOB.
DR HSSP; P00766; 1ACB.
DR MEROPS; S01.152; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser. protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.

DR InterPro; IPR001314; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen.
FT CHAIN 1 13 CHYMOTRYPSIN B, A CHAIN.
FT CHAIN 16 146 CHYMOTRYPSIN B, A CHAIN.
FT CHAIN 149 245 CHYMOTRYPSIN B, B CHAIN.
FT ACT_SITE 57 57 CHYMOTRYPSIN B, C CHAIN.
FT ACT_SITE 102 102 CHARGE RELAY SYSTEM.
FT ACT_SITE 195 195 CHARGE RELAY SYSTEM.
FT DISULFID 1 122
FT DISULFID 42 58
FT DISULFID 136 201
FT DISULFID 168 182
FT DISULFID 191 220
SQ SEQUENCE 245 AA; 25755 MW; 678016446FF5FE95 CRC64;

Query Match 82.0%; Score 82; DB 1; Length 245;
Best Local Similarity 78.9%; Pred. No. 5.9e-06;
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 IVNGEEAVPHXKXWQVSLQ 19
|||||:|:|:|
DB 16 IVNGEDAVPGSWPQVSLQ 34

RESULT 7
CTRB_CANFA
ID CTR2_CANFA STANDARD; PRT; 263 AA.
AC P04813;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymotrypsinogen 2 precursor (EC 3.4.21.1).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84170253; PubMed=6584866;
RA Pinsky S.D., Laforge K.S., Luc V., Scheele G.;
RT "Identification of cDNA clones encoding secretory isoenzyme forms:
sequence determination of canine pancreatic chymotrypsinogen 2
mRNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:7486-7490(1983).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
Phe-|-Xaa, Leu-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; K01173; AAA30841.1; -.
DR PIR; A21195; A21195.
DR HSSP; P00766; 1ACB.
DR MEROPS; S01.152; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser. protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen; Signal.
FT SIGNAL 1 18
FT CHAIN 19 263 CHYMOTRYPSINOGEN 2.

```

FT CHAIN 19 31 CHYMOTRYPSIN 2, A CHAIN.  
FT CHAIN 34 164 CHYMOTRYPSIN 2, B CHAIN.  
FT CHAIN 167 263 CHYMOTRYPSIN 2, C CHAIN.  
FT ACT SITE 75 75 CHARGE RELAY SYSTEM.  
FT ACT SITE 120 120 CHARGE RELAY SYSTEM.  
FT ACT SITE 213 213 CHARGE RELAY SYSTEM.  
FT DISULFID 19 140 BY SIMILARITY.  
FT DISULFID 60 76 BY SIMILARITY.  
FT DISULFID 154 219 BY SIMILARITY.  
FT DISULFID 186 200 BY SIMILARITY.  
FT DISULFID 209 238 BY SIMILARITY.  
SQ SEQUENCE 263 AA; 27787 MW; 2A2F49D813B3961 CRC64;  
  
Query Match 82.0%; Score 82; DB 1; Length 263;  
Best Local Similarity 78.9%; Pred. No. 6.4e-06;  
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 IVNGEAVPHXWQVSLQ 19  
|||||:|||||  
Db 34 IVNGEDAVPGSWPQVSLQ 52  
  
RESULT 8  
ID CTRB\_HUMAN STANDARD; PRT; 263 AA.  
AC P17538;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Chymotrypsinogen B precursor (EC 3.4.21.1).  
GN CTB1 OR CTRB.  
OS Homo sapiens (Human).  
OC Eukaryota; Euteleostomi; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Euteleostomi; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RX MEDLINE=89134264; PubMed=2917002;  
RA Tomita N., Isumoto Y., Horii A., Doi S., Yokouchi H., Ogawa M.,  
RA Mori T., Matsubara K.;  
RT "Molecular cloning and nucleotide sequence of human pancreatic  
prechymotrypsinogen cDNA."  
RL Biochem. Biophys. Res. Commun. 158:569-575(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RA Strauberg R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr|-Xaa, Trp|-Xaa,  
Phe|-Xaa, Leu|-Xaa.  
CC -1- SUBCELLULAR LOCATION: Extracellular.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
  
EMBL; M24400; AAA52128.1; -.  
DR EMBL; BC005385; AAH05385.1; -.  
DR PIR; A31289; A31299.  
DR HSP; P00766; 1CHG.  
DR MEROPS; S01.152; -.  
DR Gnew; HGNC:2521; CTRB1.  
DR MIM; 118890; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser. protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.

DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS02040; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal.  
FT SIGNAL 1 18  
FT CHAIN 19 263 CHYMOTRYPSIN B.  
FT CHAIN 19 31 CHYMOTRYPSIN B, A CHAIN.  
FT CHAIN 34 164 CHYMOTRYPSIN B, B CHAIN.  
FT CHAIN 167 263 CHYMOTRYPSIN B, C CHAIN.  
FT ACT SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 19 140 BY SIMILARITY.  
FT DISULFID 60 76 BY SIMILARITY.  
FT DISULFID 154 219 BY SIMILARITY.  
FT DISULFID 186 200 BY SIMILARITY.  
FT DISULFID 209 238 BY SIMILARITY.  
SQ SEQUENCE 263 AA; 27870 MW; 4C1C055A490B8701 CRC64;  
  
Query Match 82.0%; Score 82; DB 1; Length 263;  
Best Local Similarity 78.9%; Pred. No. 6.4e-06;  
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 IVNGEAVPHXWQVSLQ 19  
|||||:|||||  
Db 34 IVNGEDAVPGSWPQVSLQ 52  
  
RESULT 9  
ID CTRB\_RAT STANDARD; PRT; 263 AA.  
AC P07338;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Chymotrypsinogen B precursor (EC 3.4.21.1).  
GN CTRB1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euteleostomi; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85054881; PubMed=6209274; Valenzuela P., Craik C.S.,  
RA Bell G.I., Quinto C., Quiroga M.,  
RA Rutter W.J.;  
RT "Isolation and sequence of a rat chymotrypsin B gene."  
RL J. Biol. Chem. 259:14265-14270(1984).  
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr|-Xaa, Trp|-Xaa,  
Phe|-Xaa, Leu|-Xaa.  
CC -1- SUBCELLULAR LOCATION: Extracellular.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
  
EMBL; K02298; AAA98732.1; -.  
DR PIR; A22658; KYRTB.  
DR HSP; P00766; 1CHG.  
DR MEROPS; S01.152; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser. protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS02040; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.



```
DR PROSITE; PS00135; TRYPSIN SER; 1.
KW Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen; Signal.
FT SIGNAL 1 18
FT CHAIN 19 263 CHYMOTRYPSINOGEN B.
FT CHAIN 19 31 CHYMOTRYPSIN B, A CHAIN.
FT CHAIN 34 164 CHYMOTRYPSIN B, B CHAIN.
FT CHAIN 167 263 CHYMOTRYPSIN B, C CHAIN.
FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 19 140 BY SIMILARITY.
FT DISULFID 60 76 BY SIMILARITY.
FT DISULFID 154 219 BY SIMILARITY.
FT DISULFID 186 200 BY SIMILARITY.
FT DISULFID 209 238 BY SIMILARITY.
SQ SEQUENCE 263 AA; 27849 MW; ACAFDACF8C4DA6D CRC64;

Query Match 81.0%; Score 81; DB 1; Length 263;
Best Local Similarity 73.7%; Pred. No. 9.2e-06;
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEAVPHYXWQVSLQ 19
Db 34 IVNGDAIPGSWPMQVSLQ 52

RESULT 10
EL3A HUMAN
ID EL3A HUMAN STANDARD; PRT; 270 AA.
DC P09093; QBRW4;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Elastase IIIA precursor (EC 3.4.21.70) (Protease E).
GN ELA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Tani T., Ohsumi J., Mita K., Takiguchi Y.;
RX MEDLINE=88087253; PubMed=2826474;
RT "Identification of a novel class of elastase isozyme, human
RL pancreatic elastase III, by cDNA and genomic gene cloning.";
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Pancreas;
RX MEDLINE=89034017; PubMed=2460440;
RA Shirasu Y., Takemura K., Yoshida H., Sato Y., Iijima H.,
RA Shimada Y., Mikayama T., Ozawa T., Ikeda N., Ishida A., Tanai Y.,
RA Matsuki S., Tanaka J., Ikenaga H., Ogawa M.;
RT "Molecular cloning of complementary DNA encoding one of the human
RL pancreatic protease E isozymes.";
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: EFFICIENT PROTEASE WITH ALANINE SPECIFICITY BUT ONLY
CC -!- LITTLE ELASTOLYTIC ACTIVITY.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Ala-Xaa. Does not
CC -!- HYDROLYSE elastin.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M18700; AAA66350.1; -
DR EMBL; M18693; AAA66350.1; JOINED.
DR EMBL; M18694; AAA66350.1; JOINED.
DR EMBL; M18695; AAA66350.1; JOINED.
DR EMBL; M18696; AAA66350.1; JOINED.
DR EMBL; M18697; AAA66350.1; JOINED.
DR EMBL; M18698; AAA66350.1; JOINED.
DR EMBL; M18699; AAA66350.1; JOINED.
DR EMBL; D00306; BAA00212.1; -.
DR EMBL; BC005918; AAH05918.1; -.
DR PIR; A29934; A29934.
DR HSP; P05805; IFON.
DR MEROPS; S01.154; -.
DR Genew; HGNC:15944; ELA3A.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Tryp.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYD_SPC; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Zymogen; Signal; Glycoprotein.
FT SIGNAL 1 15
FT CHAIN 16 28 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 29 270 ELASTASE IIIA.
FT ACT_SITE 73 73 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 123 123 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 58 74 BY SIMILARITY.
FT DISULFID 117 120 PROBABLE.
FT DISULFID 157 223 BY SIMILARITY.
FT DISULFID 188 204 BY SIMILARITY.
FT DISULFID 213 244 BY SIMILARITY.
FT CARBOHYD 114 114 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 63 63 I -> T (IN REF. 3).
FT CONFLICT 106 106 MISSING (IN REF. 1).
FT CONFLICT 174 174 K -> E (IN REF. 3).
SQ SEQUENCE 270 AA; 29474 MW; 576DD255A4118C CRC64;

Query Match 80.0%; Score 80; DB 1; Length 270;
Best Local Similarity 68.4%; Pred. No. 1.4e-05;
Matches 13; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVNGEAVPHYXWQVSLQ 19
Db 29 VVHGDAVPYSWPMQVSLQ 47

RESULT 11
CLCR RAT
ID CLCR RAT STANDARD; PRT; 268 AA.
AC P55091; Q63198;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Caldecrin precursor (EC 3.4.21.2) (Chymotrypsin C) (Serum calcium-
DE decreasing factor).
GN CTRC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Pancreas;
RX MEDLINE=96107178; PubMed=8530454;
RA Tomomura A., Tomomura M., Fukushima T., Akiyama M., Kubota N.,
RA Kumaki K., Nishii Y., Noikura T., Saeki T.;
RT "Molecular cloning and expression of serum calcium-decreasing factor
```

```

RT (caldecrin).";
RL J. Biol. Chem. 270:30315-30321(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=92165057; PubMed=1537555;
RA Kang J., Wiegand U., Mueller-Hill B.;
RT "Identification of cDNAs encoding two novel rat pancreatic serine
RL proteases.";
RL Gene 110:181-187(1992).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98207038; PubMed=9538241;
RA Yoshino-Yasuda I., Kobayashi K., Akiyama M., Itoh H., Tomomura A.,
RA Saheki T.;
RT "Caldecrin is a novel-type serine protease expressed in pancreas, but
RT its homologue, elastase IV, is an artifact during cloning derived
RT from caldecrin gene.";
RL J. Biochem. 123:546-554(1998).
CC -1- FUNCTION: HAS CHYMOTRYPSIN-TYPE PROTEASE ACTIVITY AND HYPOCALCEMIC
CC ACTIVITY.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Leu|-Xaa, Tyr|-Xaa,
CC Phe|-Xaa, Met|-Xaa, Trp|-Xaa, Gln|-Xaa, Asn|-Xaa.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.
CC -1- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE ELASTASE IV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; S80379; AAB35830.1; -.
DR EMBL; X59014; CAA41753.1; -.
DR HSSP; P00766; ICHG.
DR MEROPS; S01.157; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR KW Hydrolyase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 16
FT PROPEP 17 29
FT CHAIN 30 268
FT ACT_SITE 74 74
FT ACT_SITE 123 123
FT ACT_SITE 216 216
FT DISULFID 17 141
FT DISULFID 59 75
FT DISULFID 155 222
FT DISULFID 186 202
FT DISULFID 212 243
FT CARBOHYD 25 25
FT CARBOHYD 90 90
FT CONFLICT 42 42
FT CONFLICT 96 120
FT EGGSVYAEVDITVYVHEKWLFLWN -> AEAPCTLRWTPS
FT TSMRSGTSSCGT (IN REF. 2).
FT SEQUENCE 268 AA; 29374 MW; 33B67AF34D0F8583 CRC64;
Query Match 78.0%; Score 78; DB 1; Length 268;
Best Local Similarity 68.4%; Pred. No. 2.8e-05;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 IYNGEAVPHXWQVSLQ 19
Db 30 VVGGEAVPNWQVSLQ 48

```

## RESULT 12

```

CLCR_HUMAN
ID CLCR_HUMAN STANDARD; PRT; 268 AA.
AC Q99895; Q99895; O00765;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Caldecrin precursor (EC 3.4.21.2) (Chymotrypsin C).
GN CTRC OR CLCR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT TRP-80.
RC TISSUE=Pancreas;
RX MEDLINE=96221265; PubMed=8635596;
RA Tomomura A., Akiyama M., Itoh H., Yoshino I., Tomomura M., Nishii Y.,
RA Noikura T., Saheki T.;
RT "Molecular cloning and expression of human caldecrin.";
RL FEBS Lett. 386:26-28(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Coville G.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 17-268 FROM N.A.
RC TISSUE=Pancreas;
RA Sziegoleit A.;
RT "A human pancreatic chymotrypsin: biochemical and molecular
RT characterization.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=98207038; PubMed=9538241;
RA Yoshino-Yasuda I., Kobayashi K., Akiyama M., Itoh H., Tomomura A.,
RA Saheki T.;
RT "Caldecrin is a novel-type serine protease expressed in pancreas, but
RT its homologue, elastase IV, is an artifact during cloning derived
RT from caldecrin gene.";
RL J. Biochem. 123:546-554(1998).
CC -1- FUNCTION: HAS CHYMOTRYPSIN-TYPE PROTEASE ACTIVITY AND HYPOCALCEMIC
CC ACTIVITY.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Leu|-Xaa, Tyr|-Xaa,
CC Phe|-Xaa, Met|-Xaa, Trp|-Xaa, Gln|-Xaa, Asn|-Xaa.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; S82198; AAB47104.2; ALT_SEQ.
DR EMBL; AL031283; CAB77355.1; -.
DR EMBL; Y13697; CAA74031.1; -.
DR HSSP; P00766; ICHG.
DR MEROPS; S01.157; -.
DR Genew; HGNC:2523; CTRC.
DR MTM; 601405; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.

```

```

KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
KW Polymorphism.
FT SIGNAL 1 16 POTENTIAL.
FT PROPEP 17 29 ACTIVATION PEPTIDE.
FT CHAIN 30 268 CALDECRIN.
FT ACT_SITE 74 74 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 121 121 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 216 216 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 17 141 BY SIMILARITY.
FT DISULFID 59 75 BY SIMILARITY.
FT DISULFID 155 222 BY SIMILARITY.
FT DISULFID 186 202 BY SIMILARITY.
FT DISULFID 212 243 BY SIMILARITY.
FT CARBOHYD 35 25 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 52 52 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 226 226 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 80 80 R -> W.
FT CONFLICT 16 16 /FTID=VAR 010928.
FT CONFLICT 52 52 S -> T (IN REF. 1).
FT CONFLICT 52 52 N -> D (IN REF. 3).
SQ SEQUENCE 268 AA; 29484 MW; 460BF33B4A96516F CRC64;

Query Match 77.0%; Score 77; DB 1; Length 268;
Best Local Similarity 63.4%; Pred. No. 4e-05;
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 IVNGEAVPHXWQVSLQ 19
Db 30 VVGGEARPHSWPQVSLQ 48

RESULT 13
EL2A_HUMAN STANDARD; PRT; 269 AA.
AC P08217; Q14243;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Elastase 2A precursor (EC 3.4.21.71).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87217962; PubMed=3646943;
RA Kawashima I., Tani T., Shimada K., Takiguchi Y.;
RT "Characterization of pancreatic elastase II cDNAs: two elastase II
RT mRNAs are expressed in human pancreas.";
RL DNA 6.163-172(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88107669; PubMed=3427074;
RA Fletcher T.S., Shen W.F., Largman C.;
RT "Primary structure of human pancreatic elastase 2 determined by
RT sequence analysis of the cloned mRNA.";
RL Biochemistry 26:7256-7261(1987).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=88198076; PubMed=2834346;
RA Shirasu Y., Yoshida H., Matsuki S., Takemura K., Ikeda N., Shimada Y.,
RA Ozawa T., Mikayama T., Iijima H., Ishida A., Sato Y., Tamai Y.,
RA Tanaka J., Ikenaga H.;
RT "Molecular cloning and expression in Escherichia coli of a cDNA
RT encoding human pancreatic elastase 2.";
RL J. Biochem. 102:1555-1563(1987).
RN [4]
RP SEQUENCE FROM N.A.
PP Thomas D.;
RA Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ACTS UPON ELASTIN.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Leu-Xaa, Met-Xaa
CC and Phe-Xaa. Hydrolyzes elastin.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: PANCREAS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; M16631; AAA52374.1; -.
CC EMBL; M16652; AAA52380.1; -.
CC EMBL; D00236; BAA00165.1; -.
CC EMBL; AL512883; CAC42421.1; -.
CC EMBL; BC007031; AAH07031.1; -.
CC PIR; A27432; A27432.
CC PIR; B26823; B26823.
CC HSSP; P00772; IEFG.
CC MEROPS; S01.155; -.
CC
CC InterPro; IPR001314; Chymotrypsin.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; TRYPSIN; 1.
CC PROSITE; PS50240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 16 ACTIVATION PEPTIDE.
FT PROPEP 17 28 ELASTASE 2A.
FT CHAIN 29 269 BY SIMILARITY.
FT DISULFID 58 74 BY SIMILARITY.
FT DISULFID 155 222 BY SIMILARITY.
FT DISULFID 186 202 BY SIMILARITY.
FT DISULFID 212 243 BY SIMILARITY.
FT ACT_SITE 73 73 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 121 121 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 216 216 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CONFLICT 202 202 C -> V (IN REF. 3).
SQ SEQUENCE 269 AA; 28888 MW; A2E05143E9F4987C CRC64;

Query Match 74.0%; Score 74; DB 1; Length 269;
Best Local Similarity 68.4%; Pred. No. 0.00012;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 IVNGEAVPHXWQVSLQ 19
Db 29 VVGGEARPHSWPQVSLQ 47

RESULT 14
EL2A_MOUSE STANDARD; PRT; 271 AA.
AC P05208;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Elastase 2 precursor (EC 3.4.21.71).
GN ELA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=87066713; PubMed=3641189;
RA Stevenson B.J., Hagenbuchle O., Wellauer P.K.;
RT "sequence organisation and transcriptional regulation of the mouse
RL elastase II and trypsin genes.";
RL Nucleic Acids Res. 14:8307-8330(1986).
CC -!- FUNCTION: ACTS UPON ELASTIN.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Leu-|-Xaa, Met-|-Xaa
CC and Phe-|-Xaa. Hydrolyzes elastin.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: PANCREAS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04573; CAA28242.1; -.
DR EMBL; X04576; CAA28244.1; -.
DR PIR; A25528; A25528.
DR HSSP; P00772; IELG.
DR MEROPS; S01.155; -.
DR MGD; MGI:95316; Ela2.
DR InterPro; IPR001314; Chymotrypsin.
DR PRINTS; PR00722; Chymotrypsin.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 16
FT PROPEP 17 30 ACTIVATION PEPTIDE.
FT CHAIN 31 271 ELASTASE 2.
FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 123 123 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 218 218 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 60 76 BY SIMILARITY.
FT DISULFID 157 224 BY SIMILARITY.
FT DISULFID 188 204 BY SIMILARITY.
FT DISULFID 214 245 BY SIMILARITY.
SQ SEQUENCE 271 AA; 28913 MW; FA542AE36FED3B4B CRC64;

Query Match 74.0%; Score 74; DB 1; Length 271;
Best Local Similarity 63.2%; Pred. No. 0.00012;
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSLQ 19
Db 31 VVGGEATPTWPQVSLQ 49

RESULT 15
EL2_RAT STANDARD; PRT; 271 AA.
AC P00774;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Elastase 2 precursor (EC 3.4.21.71).
GN ELA2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82182967; PubMed=6918221;
RA McDonald R.J., Swift G.H., Quinto C., Swain W., Pictet R.L.,

```

```

RA Nikovits W., Rutter W.J.;
RT "Primary structure of two distinct rat pancreatic preproelastases
RT determined by sequence analysis of the complete cloned messenger
RL ribonucleic acid sequences.";
RL Biochemistry 21:1453-1463(1982).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85054882; PubMed=6094548;
RA Swift G.H., Craik C.S., Stary S.J., Quinto C., Lahaie R.G.,
RA Rutter W.J., Macdonald R.J.;
RT "Structure of the two related elastase genes expressed in the rat
RT pancreas.";
RL J. Biol. Chem. 259:14271-14278(1984).
CC -!- FUNCTION: ACTS UPON ELASTIN.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Leu-|-Xaa, Met-|-Xaa
CC and Phe-|-Xaa. Hydrolyzes elastin.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: PANCREAS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V01233; CAA24543.1; -.
DR EMBL; L00124; AAA98780.1; -.
DR EMBL; L00118; AAA98780.1; JOINED.
DR EMBL; L00119; AAA98780.1; JOINED.
DR EMBL; L00120; AAA98780.1; JOINED.
DR EMBL; L00121; AAA98780.1; JOINED.
DR EMBL; L00122; AAA98780.1; JOINED.
DR EMBL; L00123; AAA98780.1; JOINED.
DR PIR; A00961; ELRT2.
DR HSSP; P00772; IELG.
DR MEROPS; S01.155; -.
DR InterPro; IPR001314; Chymotrypsin.
DR PRINTS; PR00722; Chymotrypsin.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 16
FT PROPEP 17 30 ACTIVATION PEPTIDE.
FT CHAIN 31 271 ELASTASE 2.
FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 123 123 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 218 218 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 60 76 BY SIMILARITY.
FT DISULFID 157 224 BY SIMILARITY.
FT DISULFID 188 204 BY SIMILARITY.
FT DISULFID 214 245 BY SIMILARITY.
SQ SEQUENCE 271 AA; 28885 MW; 125C783B857B71E3 CRC64;

Query Match 72.0%; Score 72; DB 1; Length 271;
Best Local Similarity 63.2%; Pred. No. 0.00035;
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSLQ 19
Db 31 VVGGEATPTWPQVSLQ 49

```

Search completed: February 12, 2003, 10:23:15  
Job time : 3.68657 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:16:06 ; Search time 14.1493 Seconds  
(without alignments)  
291.248 Million cell updates/sec

Title: US-10-036-371-4

Perfect score: 100

Sequence: 1 IVNGEAVPHXWQVSLQX 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21:.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacterioph.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	95	95.0	263	13 Q9PWQ6	Q9PWQ6 gadus morhu
2	94	94.0	269	11 Q9D7T9	Q9D7T9 mus musculu
3	94	94.0	269	11 Q9C522	Q9C522 mus musculu
4	92	92.0	260	13 Q9W7Q3	Q9W7Q3 paraliichthy
5	87	87.0	269	6 Q9SKW7	Q9SKW7 bos taurus
6	85	85.0	257	6 O19023	O19023 macaca mula
7	81	81.0	263	11 Q9DC86	Q9DC86 mus musculu
8	81	81.0	263	11 Q9D8X8	Q9D8X8 mus musculu
9	81	81.0	263	11 Q9CR35	Q9CR35 mus musculu
10	80	80.0	264	11 Q9D960	Q9D960 mus musculu
11	80	80.0	264	11 Q9D7F8	Q9D7F8 mus musculu
12	80	80.0	264	11 Q9EQ28	Q9EQ28 rattus norv
13	80	80.0	264	11 Q9ER05	Q9ER05 mus musculu
14	80	80.0	270	4 Q96QL8	Q96QL8 homo sapien
15	77	77.0	461	5 Q8T4N2	Q8T4N2 rhipicephal
16	74	74.0	40	13 Q9PSP2	Q9PSP2 gallus gall

17	71	71.0	261	13 Q9W7Q4	Q9W7Q4 paraliichthy
18	69	69.0	20	13 Q9PRR4	Q9PRR4 scyliorhinu
19	69	69.0	269	4 Q96QV5	Q96QV5 homo sapien
20	69	69.0	270	13 Q91039	Q91039 gadus morhu
21	69	69.0	474	5 Q8T4N3	Q8T4N3 rhipicephal
22	68	68.0	20	13 Q9PRR3	Q9PRR3 scyliorhinu
23	67	67.0	1004	13 P79953	P79953 xenopus lae
24	66	66.0	249	13 Q9W7Q1	Q9W7Q1 paraliichthy
25	66	66.0	266	13 Q9W7Q0	Q9W7Q0 paraliichthy
26	66	66.0	268	13 Q9W7Q2	Q9W7Q2 paraliichthy
27	66	66.0	492	4 Q96T73	Q96T73 homo sapien
28	65	65.0	1524	13 Q91674	Q91674 xenopus lae
29	63	63.0	20	13 Q9PRT8	Q9PRT8 gadus morhu
30	63	63.0	267	5 Q9BK47	Q9BK47 luidia foll
31	62	62.0	265	5 O18488	O18488 penaeus van
32	62	62.0	270	5 Q8WR11	Q8WR11 paraliithode
33	62	62.0	405	4 Q96E86	Q96E86 homo sapien
34	62	62.0	806	6 O18783	O18783 macropus eu
35	61	61.0	266	13 Q92077	Q92077 gadus morhu
36	61	61.0	417	11 Q8VHK8	Q8VHK8 mus musculu
37	61	61.0	417	11 Q8VDV1	Q8VDV1 mus musculu
38	61	61.0	810	4 Q15146	Q15146 homo sapien
39	60	60.0	812	11 Q9R0W3	Q9R0W3 rattus norv
40	60	60.0	812	11 Q91WJ5	Q91WJ5 mus musculu
41	59	59.0	264	13 Q8QGf6	Q8QGf6 xenopus lae
42	59	59.0	268	5 Q17030	Q17030 anopheles g
43	59	59.0	271	5 Q8T9R6	Q8T9R6 culix pipie
44	58	58.0	270	5 Q96871	Q96871 trichinella
45	58	58.0	273	11 Q921N4	Q921N4 mus musculu

## ALIGNMENTS

## RESULT 1

Q9PWQ6 PRELIMINARY; PRT; 263 AA.  
ID Q9PWQ6  
AC Q9PWQ6  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Chymotrypsin B precursor (EC 3.4.21.1).  
GN CHXB  
OS Gadus morhua (Atlantic cod).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.  
OX NCBI\_TaxID=8049;  
RN [1]  
RC SEQUENCE FROM N.A.  
RC TISSUE=PYLORIC CAECA;  
RX MEDLINE=20464334; PubMed=11011764;  
RA Spilliaert R.; Gudmundsdottir A.;  
RT "Molecular Cloning of the Atlantic Cod Chymotrypsinogen B.";  
RL Microb. Comp. Genomics 5:41-50(2000).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.  
CC EMBL; AJ242521; CAB43766.1; --  
DR HSSP; P00766; 1CHG.  
DR MEROPS; S01.152; --  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser.protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SRR; 1.  
KW Hydrolase; Serine protease; Signal  
FT SIGNAL 1 16 POTENTIAL.  
FT CHAIN 17 263 CHYMOTRYPSIN B.  
SQ SEQUENCE 263 AA; 28175 MW; EF61B18A34EE5E7C CRC64;

```

Query Match      95.0%; Score 95; DB 13; Length 263;
Best Local Similarity 89.5%; Pred. No. 4.2e-08;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSLQ 19
DB 32 IVNGEEAVPHSWPQVSLQ 50

RESULT 2
Q9D7T9 PRELIMINARY; PRT; 269 AA.
ID Q9D7T9
AC Q9D7T9
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 21, Last annotation update)
DE 2310074F01RIK protein.
GN ELA3B OR 2310074F01RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=STOMACH;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC EMBL; AK010149; BAB26734.1; -.
DR EMBL; AK009129; BAB26092.1; -.
DR HSSP; P05805; 1FON.
DR MEROPS; S01.154; -.
DR MGD; MGI:1915118; Ela3b.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 269 AA; 28977 MW; 9F43F769DDB2A7CF CRC64;

Query Match      94.0%; Score 94; DB 11; Length 269;
Best Local Similarity 84.2%; Pred. No. 6.3e-08;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSLQ 19
DB 28 VVNGEEAVPHSWPQVSLQ 46

RESULT 4
Q9W7Q3 PRELIMINARY; PRT; 260 AA.
ID Q9W7Q3
AC Q9W7Q3
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Chymotrypsinogen 2.
OS Paralicthys olivaceus (Flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Schriml L.M., Staubli F., Su  
RA Sakai K., Okida T., Furuno M.

RA	Schr
RA	Saka

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Guastincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.  
CC ENBL; AK003060; BAB22539.1; -.  
DR HSSP; P00766; 1GCT.  
DR MEROPS; S01.152; -.  
DR MGD; MGI:1913723; 2200008D09Rik.  
DR InterPro; IPR001314; Chymotrypsin.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease.  
SQ SEQUENCE 263 AA; 27821 MW; 2620A27AFBA5D04D CRC64;  
  
Query Match 81.0%; Score 81; DB 11; Length 263;  
Best Local Similarity 73.7%; Pred. No. 9.9e-06;  
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 IVNGEAVPHXWQVSLQ 19  
|||||:|:|:|:|:|  
Db 34 IVNGEDAIPGWPWQVSLQ 52  
  
RESULT 8  
ID Q9DBX8 PRELIMINARY; PRT; 263 AA.  
AC Q9DBX8;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE 2200008D09Rik protein.  
DE 2200008D09Rik.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=STOMACH, SPLEEN, AND PANCREAS;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Giasi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Kamiya M., Lee N.H.,  
RA Guastincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.  
CC ENBL; AK008927; BAB25971.1; -.  
DR EMBL; AK003079; BAB22553.1; -.  
DR EMBL; AK007765; BAB25241.1; -.  
DR EMBL; AK007815; BAB25280.1; -.  
DR EMBL; AK008729; BAB25861.1; -.  
DR EMBL; AK008888; BAB25954.1; -.  
DR HSSP; P00766; 1GCT.

RL Nature 409:685-690(2001).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.  
CC ENBL; AK007566; BAB25112.1; -.  
DR HSSP; P00766; 1GCT.  
DR MEROPS; S01.152; -.  
DR MGD; MGI:1913723; 2200008D09Rik.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease.  
SQ SEQUENCE 263 AA; 27898 MW; C0638FB8F905A92F CRC64;  
  
Query Match 81.0%; Score 81; DB 11; Length 263;  
Best Local Similarity 73.7%; Pred. No. 9.9e-06;  
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 IVNGEAVPHXWQVSLQ 19  
|||||:|:|:|:|:|  
Db 34 IVNGEDAIPGWPWQVSLQ 52  
  
RESULT 9  
ID Q9CR35 PRELIMINARY; PRT; 263 AA.  
AC Q9CR35;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE 2200008D09Rik protein.  
DE 2200008D09Rik.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=STOMACH, SPLEEN, AND PANCREAS;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Giasi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Kamiya M., Lee N.H.,  
RA Guastincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.  
CC ENBL; AK008927; BAB25971.1; -.  
DR EMBL; AK003079; BAB22553.1; -.  
DR EMBL; AK007765; BAB25241.1; -.  
DR EMBL; AK007815; BAB25280.1; -.  
DR EMBL; AK008729; BAB25861.1; -.  
DR EMBL; AK008888; BAB25954.1; -.  
DR HSSP; P00766; 1GCT.



```

DR MEROPS; S01.152; -.
DR MGD; MGI:1913723; 2200008D098IK.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser.protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease.
SQ SEQUENCE 263 AA; 27822 MW; 28C4487AF1A26B27 CRC64;

Query Match      81.0%; Score 81; DB 11; Length 263;
Best Local Similarity 73.7%; Pred. No. 9.9e-06;
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEAVPHXKXWQVSLQ 19
   |||||:|:|:|
Db 34 IVNGDAIRGSPWQVSLQ 52

RESULT 10
Q9D960 PRELIMINARY; PRT; 264 AA.
AC Q9D960;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 1810004D15RIK protein.
GN CTRL OR 1810004D15RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Carninci P., de Bonaldo M.F.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayaishizaki Y.;
RT "Functional annotation of a full-length mouse cdna collection.";
RL Nature 409:685-690(2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AK007333; BAB24967.1; -.
DR HSSP; P00766; 4CHA.
DR MEROPS; S01.256; -.
DR MGD; MGI:88558; Ctrl.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser.protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease.
SQ SEQUENCE 264 AA; 28167 MW; 1D979469A07056C2 CRC64;

```

```

KW Hydrolyase; Serine protease.
SQ SEQUENCE 264 AA; 28151 MW; 1D979719E07C16DE CRC64;

Query Match      80.0%; Score 80; DB 11; Length 264;
Best Local Similarity 78.9%; Pred. No. 1.5e-05;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 IVNGEAVPHXKXWQVSLQ 19
   |||||:|:|:|
Db 34 IVNGENAVPGSPWQVSLQ 52

RESULT 11
Q9D7P8 PRELIMINARY; PRT; 264 AA.
AC Q9D7P8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 1810004D15RIK protein.
GN CTRL OR 1810004D15RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Carninci P., de Bonaldo M.F.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayaishizaki Y.;
RT "Functional annotation of a full-length mouse cdna collection.";
RL Nature 409:685-690(2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AK009019; BAB26029.1; -.
DR HSSP; P00766; 4CHA.
DR MEROPS; S01.256; -.
DR MGD; MGI:88558; Ctrl.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser.protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease.
SQ SEQUENCE 264 AA; 28167 MW; 1D979469A07056C2 CRC64;

Query Match      80.0%; Score 80; DB 11; Length 264;
Best Local Similarity 78.9%; Pred. No. 1.5e-05;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 IVNGEAVPHXKXWQVSLQ 19
   |||||:|:|:|
Db 34 IVNGENAVPGSPWQVSLQ 52

```

Wed Feb 12 11:59:39 2003

```

DR HSP; P00766; 4CHA.
DR MEROPS; S01.256; -.
DR MGD; MGI:88558; Ctrl.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin_1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease.
KW SEQUENCE 264 AA; 28135 MW; 1D979709A07056C2 CRC64;
SQ
Query Match 80.0%; Score 80; DB 11; Length 264;
Best Local Similarity 78.9%; Pred. No. 1.5e-05;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 IVNGEAVPHXWQVSLQ 19
Db 34 IVNGENAVGSPWQVSLQ 52
RESULT 14
Q96QL8 PRELIMINARY; PRT; 270 AA.
ID Q96QL8
AC Q96QL8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to elastase 3, pancreatic (protease E).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC008383; AAH08383.1; -.
DR MEROPS; S01.154; -.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin_1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 270 AA; 29446 MW; 946DDBDA694A102E CRC64;
Query Match 80.0%; Score 80; DB 4; Length 270;
Best Local Similarity 68.4%; Pred. No. 1.5e-05;
Matches 13; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 1 IVNGEAVPHXWQVSLQ 19
Db 29 VVHGEDAVPYSPWQVSLQ 47
RESULT 15
Q8T4N2 PRELIMINARY; PRT; 461 AA.
ID Q8T4N2
AC Q8T4N2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Midgut serine proteinase-3.
OS Rhipicephalus appendiculatus (Brown ear tick).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Parasitiformes; Ixodida; Ixodidae; Rhipicephalus.
OX NCBI_TaxID=34631;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RESULT 12
Q9EQZ8 PRELIMINARY; PRT; 264 AA.
ID Q9EQZ8
AC Q9EQZ8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Chymopasin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RAT PANCREAS;
RA Sogame Y., Mitsui S., Kataoka K., Kashima K., Kato M., Sakagami J.,
RA Yamaguchi N.;
RL "Molecular cloning of rat chymopasin."
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.
DR EMBL; AB020757; BAB20287.1; -.
DR HSP; P00766; 4CHA.
DR MEROPS; S01.256; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin_1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 264 AA; 28116 MW; F9ED5D210FD3500E CRC64;
Query Match 80.0%; Score 80; DB 11; Length 264;
Best Local Similarity 78.9%; Pred. No. 1.5e-05;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 IVNGEAVPHXWQVSLQ 19
Db 34 IVNGENAVGSPWQVSLQ 52
RESULT 13
Q9ER05 PRELIMINARY; PRT; 264 AA.
ID Q9ER05
AC Q9ER05;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Chymopasin (Chymotrypsin A CTRA-1).
GN CTRL OR CTRA1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Mitsui S., Yamaguchi N.;
RL "Molecular cloning of mouse chymopasin."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129S6/SVEVITAC; TISSUE=SPLEEN;
RA Bjoernlett M.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.
DR EMBL; AB016228; BAB20275.1; -.
DR EMBL; AF236365; AAL11034.1; -.

```

RA Mulenga A., Onuma M., Sugimoto C.;  
RT "Rhipicephalus appendiculatus midgut serine proteinase-3, cDNA cloning  
and characterization";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY078095; AAL79567.1; ...  
SQ SEQUENCE 461 AA; 49581 MW; 641C3DCA5CF2EB68 CRC64;

Query Match 77.0%; Score 77; DB 5; Length 461;  
Best Local Similarity 68.4%; Pred. NO. 8.6e-05;  
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSLQ 19  
|||:|||||:|:|  
Db 208 IVAGQEAIPHSPWQASVQ 226

Search completed: February 12, 2003, 10:27:23  
Job time : 15.1493 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:18:40 ; Search time 4.29851 Seconds  
(without alignments)  
136.898 Million cell updates/sec

Title: US-10-036-371-4

Perfect score: 100

Sequence: 1 INVGEAVPHYKXWQVSLQX 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfilees1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85	85.0	31	5	PCT-US95-16826-1
2	85	85.0	32	5	PCT-US95-16826-2
3	85	85.0	228	1	US-08-278-091-10
4	85	85.0	228	1	US-08-483-859-10
5	85	85.0	228	1	US-08-483-859-10
6	85	85.0	228	1	US-08-483-859-10
7	85	85.0	228	2	US-08-487-167-10
8	85	85.0	228	2	US-08-482-816-10
9	85	85.0	228	2	US-08-296-149-10
10	85	85.0	228	2	US-08-801-499-10
11	85	85.0	228	3	US-08-615-271-10
12	85	85.0	228	3	US-09-074-660-10
13	85	85.0	228	3	US-09-074-659-10
14	85	85.0	228	4	US-09-106-468-10
15	85	85.0	228	4	US-09-106-468-10
16	85	85.0	228	4	US-09-106-467-10
17	82	82.0	242	4	US-08-944-483-57
18	82	82.0	20	2	US-08-385-540A-16
19	82	82.0	20	3	US-08-600-273A-16
20	82	82.0	20	3	US-08-486-820-16
21	82	82.0	229	2	US-08-557-146-13
22	82	82.0	229	2	US-09-154-344-13
23	82	82.0	230	4	US-08-944-483-62
24	82	82.0	231	2	US-09-027-337-6
25	82	82.0	231	4	US-09-644-600-6
26	80	80.0	242	4	US-08-944-483-58
27	77	77.0	239	4	US-08-944-483-61

28	77	77.0	268	1	US-08-270-584A-2
29	77	77.0	268	1	US-08-568-031-2
30	77	77.0	268	2	US-08-765-192-2
31	77	77.0	268	2	US-08-966-319-2
32	77	77.0	268	3	US-08-153-304-2
33	77	77.0	268	3	US-08-199-793-2
34	74	74.0	241	4	US-08-944-483-59
35	69	69.0	241	4	US-08-944-483-60
36	66	66.0	283	3	US-08-807-151-1
37	66	66.0	283	4	US-09-478-957-1
38	66	66.0	492	4	US-09-342-749-2
39	66	66.0	492	4	US-09-691-840-2
40	62	62.0	213	3	US-08-906-769-149
41	62	62.0	213	3	US-08-906-616-149
42	62	62.0	213	4	US-08-639-075A-149
43	62	62.0	213	4	US-09-012-431-149
44	62	62.0	213	4	US-09-012-692-149
45	62	62.0	213	4	US-08-906-613-149

## ALIGNMENTS

## RESULT 1

PCT-US95-16826-1

; Sequence 1, Application PC/TUS9516826

; GENERAL INFORMATION:

; APPLICANT:

; APPLICANT: NAME: OSTEOSA, INC.

; APPLICANT: STREET: 2040 Babcock Road, Suite 201

; APPLICANT: CITY: San Antonio

; APPLICANT: STATE: Texas

; APPLICANT: COUNTRY: USA

; APPLICANT: POSTAL CODE: 78229

; APPLICANT: NAME: YONEDA, Tohiyuki

; APPLICANT: STREET: 3530 Hunter's Sound

; APPLICANT: CITY: San Antonio

; APPLICANT: STATE: Texas

; APPLICANT: COUNTRY: USA

; APPLICANT: POSTAL CODE: 78230

; APPLICANT: NAME: IZBICKA, Elzbieta

; APPLICANT: STREET: 7738 Apple Green

; APPLICANT: CITY: San Antonio

; APPLICANT: STATE: Texas

; APPLICANT: POSTAL CODE: 78240

; APPLICANT: NAME: MUNDY, Gregory R.

; APPLICANT: STREET: 3719 Morgan's Creek

; APPLICANT: CITY: San Antonio

; APPLICANT: STATE: Texas

; APPLICANT: COUNTRY: USA

; APPLICANT: POSTAL CODE: 78230

; TITLE OF INVENTION: Modulators of Bone Cell Function and

; TITLE OF INVENTION: Uses Thereof

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Rhone-Poulenc Rorer Inc.

; STREET: 500 Arcola Rd. 3C43

; CITY: Collegeville

; STATE: PA

; COUNTRY: USA

; ZIP: 19002

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/16826

; FILING DATE:

•

TELEPHONE: (416) 595-1155  
 TELEFAX: (416) 595-1163  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 228 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-278-091-10

Query Match 85.0%; Score 85; DB 1; Length 228;  
 Best Local Similarity 84.2%; Pred. No. 2.5e-07;  
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXKXWQVSLQ 19  
 ||||| |  
 Db 1 IVNGEEAVPGSWPQVSLQ 19

## RESULT 4

US-08-483-859-10  
 ; Sequence 10, Application US/08483859  
 ; Patent No. 5656436  
 ; GENERAL INFORMATION:

APPLICANT: LOOSMORE, Sheena M.  
 APPLICANT: YANG, Yan-Ping  
 APPLICANT: CHONG, Pele  
 APPLICANT: COHEN, Raymond P.  
 APPLICANT: KLEIN, Michel H.  
 TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with  
 NUMBER OF SEQUENCES: 23  
 NUMBER OF SEQUENCES: 23  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney  
 STREET: Suite 701, 330 University Avenue  
 CITY: Toronto  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: M5G 1R7

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/483,859  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/296,149  
 FILING DATE: 26-AUG-1994

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/278,091  
 FILING DATE: 21-JUL-1994

ATTORNEY/AGENT INFORMATION:  
 NAME: Stewart, Michael I.

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-495 MIS.vg

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 228 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-483-859-10

Query Match 85.0%; Score 85; DB 1; Length 228;  
 Best Local Similarity 84.2%; Pred. No. 2.5e-07;  
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXKXWQVSLQ 19  
 ||||| |  
 Db 1 IVNGEEAVPGSWPQVSLQ 19

## RESULT 5

US-08-472-173-10  
 ; Sequence 10, Application US/08472173  
 ; Patent No. 5665353  
 ; GENERAL INFORMATION:

APPLICANT: LOOSMORE, Sheena M.  
 APPLICANT: YANG, Yan-Ping  
 APPLICANT: CHONG, Pele  
 APPLICANT: COHEN, Raymond P.  
 APPLICANT: KLEIN, Michel H.  
 TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with  
 NUMBER OF SEQUENCES: 23  
 NUMBER OF SEQUENCES: 23  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney  
 STREET: Suite 701, 330 University Avenue  
 CITY: Toronto  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: M5G 1R7

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/472,173  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/296,149  
 FILING DATE: 26-AUG-1994

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/278,091  
 FILING DATE: 21-JUL-1994

ATTORNEY/AGENT INFORMATION:  
 NAME: Stewart, Michael I.

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-493 MIS.vg

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 228 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-472-173-10

Query Match 85.0%; Score 85; DB 1; Length 228;  
 Best Local Similarity 84.2%; Pred. No. 2.5e-07;  
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXKXWQVSLQ 19  
 ||||| |  
 Db 1 IVNGEEAVPGSWPQVSLQ 19

## RESULT 6

US-08-487-167-10  
 ; Sequence 10, Application US/08487167  
 ; Patent No. 5869302  
 ; GENERAL INFORMATION:

APPLICANT: LOOSMORE, Sheena M.  
 APPLICANT: YANG, Yan-Ping  
 APPLICANT: CHONG, Pele  
 APPLICANT: COHEN, Raymond P.

us-10-036-371-4.ra1

Wed Feb 12 11:59:37 2003

```

;
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
;   Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,167
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
;
; QUERY MATCH
; Best Local Similarity 85.0%; Score 85; DB 2; Length 228;
; Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
;
; QY 1 IVNGEAVPHXKXWQVSLQ 19
; DB 1 IVNGEAVPGSPWQVSLQ 19
;
; RESULT 7
; US-08-482-816-10
; Sequence 10, Application US/08482816
; Patent No. 5935573
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
;   Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

;
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 08/482,816
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
;
; QUERY MATCH
; Best Local Similarity 85.0%; Score 85; DB 2; Length 228;
; Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
;
; QY 1 IVNGEAVPHXKXWQVSLQ 19
; DB 1 IVNGEAVPGSPWQVSLQ 19
;
; RESULT 8
; US-08-296-149-10
; Sequence 10, Application US/08296149
; Patent No. 5939297
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
;   Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-390
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1153
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-482-816-10

```



; LENGTH: 228 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 US-08-296-149-10

Query Match 85.0%; Score 85; DB 2; Length 228;  
 Best Local Similarity 84.2%; Pred. No. 2.5e-07;  
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 IVNGEAVPHXWQVSLQ 19  
 ||||| |  
 Db 1 IVNGEAVPGSWPWQVSLQ 19

## RESULT 9

US-08-801-499-10  
 ; Sequence 10, Application US/08801499  
 ; Patent No. 5962430

; GENERAL INFORMATION:  
 ; APPLICANT: LOOSMORE, Sheena M

; APPLICANT: YANG, Yan-Ping  
 ; APPLICANT: CHONG, Pele

; APPLICANT: OOMEN, Raymond P.  
 ; APPLICANT: KLEIN, Michel H.

; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced Protease Act

; NUMBER OF SEQUENCES: 23  
 ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney  
 ; STREET: 6th Floor, 330 University Avenue

; CITY: Toronto  
 ; STATE: Ontario

; COUNTRY: Canada  
 ; ZIP: M5G 1R7

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/801,499  
 ; FILING DATE: 21-JUL-1994

; CLASSIFICATION: 514  
 ; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/482,816  
 ; FILING DATE: 07-JUN-1995

; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/296,149

; FILING DATE: 26-AUG-1994  
 ; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/278,091  
 ; FILING DATE: 21-JUL-1994

; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Stewart, Michael I

; REGISTRATION NUMBER: 24,973  
 ; REFERENCE/DOCKET NUMBER: 1038-671 MTS:jb

; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1163  
 ; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 228 amino acids

; TYPE: amino acid  
 ; STRANDEDNESS: single

; TOPOLOGY: linear  
 US-08-801-499-10

Query Match 85.0%; Score 85; DB 2; Length 228;  
 Best Local Similarity 84.2%; Pred. No. 2.5e-07;  
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 IVNGEAVPHXWQVSLQ 19  
 ||||| |

Db 1 IVNGEAVPGSWPWQVSLQ 19

## RESULT 10

US-08-615-271-10  
 ; Sequence 10, Application US/08615271  
 ; Patent No. 5981503

; GENERAL INFORMATION:  
 ; APPLICANT: LOOSMORE, Sheena M.

; APPLICANT: YANG, Yan-Ping  
 ; APPLICANT: CHONG, Pele

; APPLICANT: OOMEN, Raymond P.  
 ; APPLICANT: KLEIN, Michel H.

; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED

; NUMBER OF SEQUENCES: 23  
 ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney  
 ; STREET: 6th Floor, 330 University Avenue

; CITY: Toronto  
 ; STATE: Ontario

; COUNTRY: Canada  
 ; ZIP: M5G 1R7

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/615,271  
 ; FILING DATE: 20-JUN-1996

; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:

; NAME: Stewart, Michael I  
 ; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-580  
 ; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1155  
 ; TELEFAX: (416) 595-1163

; INFORMATION FOR SEQ ID NO: 10:  
 ; SEQUENCE CHARACTERISTICS:

; LENGTH: 228 amino acids  
 ; TYPE: amino acid

; STRANDEDNESS: single  
 ; TOPOLOGY: linear

US-08-615-271-10

Query Match 85.0%; Score 85; DB 2; Length 228;  
 Best Local Similarity 84.2%; Pred. No. 2.5e-07;  
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 IVNGEAVPHXWQVSLQ 19  
 ||||| |

Db 1 IVNGEAVPGSWPWQVSLQ 19

## RESULT 11

US-09-074-660-10  
 ; Sequence 10, Application US/09074660  
 ; Patent No. 6020183

; GENERAL INFORMATION:  
 ; APPLICANT: LOOSMORE, Sheena M.

; APPLICANT: YANG, Yan-Ping  
 ; APPLICANT: CHONG, Pele

; APPLICANT: OOMEN, Raymond P.  
 ; APPLICANT: KLEIN, Michel H.

; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with

; NUMBER OF SEQUENCES: 23  
 ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney  
 ; STREET: 6th Floor, 330 University Avenue

; CITY: Toronto

```

CLASSIFICATION:
PRIOR APPLICATION DATA: US 08/487,167
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/278,091
FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-730 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 228 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-074-659-10

Query Match 85.0%; Score 85; DB 3; Length 228;
Best Local Similarity 84.2%; Pred. NO. 2.5e-07;
Matches 16; Conservative 0; Mismatches 3; Indels 0

QY 1 IVNGEEAVPHYXWQVSLQ 19
||||||| - |||||
Db 1 IVNGEEAVPGSPWQVSLQ 19

RESULT 13
US-09-106-468-10
; Sequence 10, Application US/09106468
; Patent No. 6114125
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
; PROTEASE ACTIVITY
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/106,468
; APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/615,271
; FILING DATE: 20-JUN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-825
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
;

```

TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 228 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-106-468-10

Query Match 85.0%; Score 85; DB 3; Length 228;  
Best Local Similarity 84.2%; Pred. No. 2.5e-07;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSLQ 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 IVNGEEAVPGSWPQVSLQ 19

RESULT 14  
US-09-106-466A-10  
; Sequence 10, Application US/09106466A  
; Patent No. 6147057  
; GENERAL INFORMATION:  
; APPLICANT: LOOSMORE, Sheena M.  
; APPLICANT: YANG, Yan-Ping  
; APPLICANT: CHONG, Pele  
; APPLICANT: COMEN, Raymond P.  
; APPLICANT: KLEIN, Michel H.  
; TITLE OF INVENTION: ANALOG OF HARMOPHILUS HIN47 WITH REDUCED  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/106.466A  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/615,271  
; FILING DATE: 20-JUN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-826  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 228 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-106-466A-10

Query Match 85.0%; Score 85; DB 4; Length 228;  
Best Local Similarity 84.2%; Pred. No. 2.5e-07;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSLQ 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 IVNGEEAVPGSWPQVSLQ 19

RESULT 15  
US-09-106-467-10  
; Sequence 10, Application US/09106467  
; Patent No. 6153580  
; GENERAL INFORMATION:  
; APPLICANT: LOOSMORE, Sheena M.  
; APPLICANT: YANG, Yan-Ping  
; APPLICANT: CHONG, Pele  
; APPLICANT: COMEN, Raymond P.  
; APPLICANT: KLEIN, Michel H.  
; TITLE OF INVENTION: ANALOG OF HARMOPHILUS HIN47 WITH REDUCED  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/106,467  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/615,271  
; FILING DATE: 20-JUN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-824  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 228 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-106-467-10

Query Match 85.0%; Score 85; DB 4; Length 228;  
Best Local Similarity 84.2%; Pred. No. 2.5e-07;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSLQ 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 IVNGEEAVPGSWPQVSLQ 19

Search completed: February 12, 2003, 10:30:09  
Job time : 5.29851 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:22:36 ; Search time 3.04478 Seconds  
(without alignments)  
167.821 Million cell updates/sec

Title: US-10-036-371-4

Perfect score: 100

Sequence: 1 IVNGEAVPHXWQVSLQX 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Score	Match	Length	ID	Description
1	95.0	20	12	US-10-036-371-4	Sequence 4, Appli
2	85.0	20	12	US-10-036-371-6	Sequence 6, Appli
3	82.0	146	10	US-09-925-297-643	Sequence 643, App
4	82.0	263	10	US-09-888-615-96	Sequence 96, Appl
5	80.0	270	10	US-09-923-779-152	Sequence 152, App
6	78.0	273	10	US-09-925-297-695	Sequence 695, App
7	74.0	269	10	US-09-925-297-576	Sequence 576, App
8	70.0	192	10	US-09-925-297-529	Sequence 529, App
9	70.0	970	10	US-09-888-615-101	Sequence 101, App
10	69.0	113	10	US-09-925-297-871	Sequence 871, App
11	66.0	283	9	US-09-981-353-23	Sequence 1, Appli
12	66.0	384	9	US-09-981-353-23	Sequence 2, Appli
13	66.0	393	9	US-10-012-896-934	Sequence 934, App
14	66.0	393	9	US-09-895-793-934	Sequence 934, App
15	66.0	393	9	US-09-895-793-934	Sequence 934, App
16	66.0	393	9	US-09-895-814-934	Sequence 934, App
17	66.0	393	10	US-09-759-143-934	Sequence 934, App
18	66.0	393	10	US-09-780-669-934	Sequence 934, App
19	66.0	492	9	US-10-012-896-895	Sequence 895, App

20	66.0	492	9	US-10-012-896-932	Sequence 932, App
21	66.0	492	9	US-09-895-793-895	Sequence 895, App
22	66.0	492	9	US-09-895-793-932	Sequence 932, App
23	66.0	492	9	US-09-895-814-895	Sequence 895, App
24	66.0	492	9	US-09-895-814-932	Sequence 932, App
25	66.0	492	10	US-09-759-143-895	Sequence 895, App
26	66.0	492	10	US-09-759-143-932	Sequence 932, App
27	66.0	492	10	US-09-780-669-895	Sequence 895, App
28	66.0	492	10	US-09-780-669-932	Sequence 932, App
29	66.0	492	10	US-09-822-827-14	Sequence 14, Appl
30	66.0	492	10	US-09-822-827-895	Sequence 895, App
31	66.0	492	10	US-09-822-827-932	Sequence 932, App
32	62.0	406	10	US-09-851-588-6	Sequence 6, Appli
33	62.0	432	9	US-10-063-547-112	Sequence 112, App
34	62.0	432	9	US-10-174-590-330	Sequence 330, App
35	62.0	432	9	US-10-176-758-330	Sequence 330, App
36	62.0	432	9	US-10-063-616-112	Sequence 112, App
37	62.0	432	9	US-10-175-737-330	Sequence 330, App
38	62.0	432	9	US-10-173-706-330	Sequence 330, App
39	62.0	432	9	US-10-173-706-330	Sequence 330, App
40	62.0	432	9	US-10-175-738-330	Sequence 330, App
41	62.0	432	9	US-10-175-732-330	Sequence 330, App
42	62.0	432	9	US-10-176-482-330	Sequence 330, App
43	62.0	432	9	US-10-176-757-330	Sequence 330, App
44	62.0	432	9	US-10-176-913-330	Sequence 330, App
45	62.0	432	9	US-10-180-552-330	Sequence 330, App

#### ALIGNMENTS

RESULT 1  
US-10-036-371-4  
; Sequence 4, Application US/10036371  
; Patent No. US20020141987A1  
; GENERAL INFORMATION:  
; APPLICANT: BJARNARSON, JON B.  
; TITLE OF INVENTION: FISH SERINE PROTEINASES AND THEIR PHARMACEUTICAL AND  
; FILE REFERENCE: COSMETIC USE  
; FILE REFERENCE: 81691/284960  
; CURRENT APPLICATION NUMBER: US/10/036,371  
; CURRENT FILING DATE: 2002-01-07  
; PRIOR APPLICATION NUMBER: 09/411,688  
; PRIOR FILING DATE: 1999-10-12  
; PRIOR APPLICATION NUMBER: 5086/99  
; PRIOR FILING DATE: 1999-06-18  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Gadus sp.  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (11)  
; OTHER INFORMATION: S or T  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (13)  
; OTHER INFORMATION: S, P or Y  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (20)  
; OTHER INFORMATION: D or Q  
US-10-036-371-4

Query Match 95.0%; Score 95; DB 12; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.7e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVNGEAVPHXWQVSLQ 19  
DB 1 IVNGEAVPHXWQVSLQ 19

Query Match 80.0%; Score 80; DB 10; Length 270;  
Best Local Similarity 68.4%; Pred. No. 9.8e-06;  
Matches 13; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSLQ 19  
:|||||:|||||  
Db 29 VVHGDAVPYSPWQVSLQ 47

## RESULT 6

US-09-925-297-695  
; Sequence 695, Application US/09925297  
; Patent No. US20020081659A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA105  
; CURRENT APPLICATION NUMBER: US/09/925,297  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05989  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 695  
; LENGTH: 273  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (27)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (28)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (34)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-297-695

Query Match 78.0%; Score 78; DB 10; Length 273;  
Best Local Similarity 68.4%; Pred. No. 2.1e-05;  
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSLQ 19  
:|||||:|||||  
Db 32 VVHGDAVPYSPWQVSLQ 50

## RESULT 7

US-09-925-297-576  
; Sequence 576, Application US/09925297  
; Patent No. US20020081659A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA105  
; CURRENT APPLICATION NUMBER: US/09/925,297  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05989  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 576  
; LENGTH: 269  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (8)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (167)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: SITE  
; LOCATION: (213)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (220)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (234)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-297-576

Query Match 74.0%; Score 74; DB 10; Length 269;  
Best Local Similarity 68.4%; Pred. No. 9e-05;  
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSLQ 19  
:|||||:|||||  
Db 34 VVGGEEARPNSWPQVSLQ 52

## RESULT 8

US-09-925-297-529  
; Sequence 529, Application US/09925297  
; Patent No. US20020081659A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA105  
; CURRENT APPLICATION NUMBER: US/09/925,297  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05989  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 529  
; LENGTH: 192  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-297-529

Query Match 70.0%; Score 70; DB 10; Length 192;  
Best Local Similarity 73.7%; Pred. No. 0.00028;  
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSLQ 19  
:|||||:|||||  
Db 30 IVNGENAVLGSWPQVSLQ 48

## RESULT 9

US-09-888-615-101  
; Sequence 101, Application US/09888615  
; Patent No. US20020084856A1  
; GENERAL INFORMATION:  
; APPLICANT: PLOWMAN, GREGORY  
; APPLICANT: WHYTE, DAVID  
; APPLICANT: CAENEPEEL, SEAN  
; APPLICANT: CHARYDCZAK, GLEN  
; APPLICANT: MANNING, GERARD  
; APPLICANT: SUDARSANAM, SUCHA  
; TITLE OF INVENTION: NOVEL PROTEASES  
; FILE REFERENCE: 038602/1214  
; CURRENT APPLICATION NUMBER: US/09/888,615  
; CURRENT FILING DATE: 2001-06-26  
; PRIOR APPLICATION NUMBER: 60/214,047  
; PRIOR FILING DATE: 2000-06-26  
; NUMBER OF SEQ ID NOS: 150  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 101  
; LENGTH: 970  
; TYPE: PRT

```
; ORGANISM: Homo sapiens
US-09-888-615-101

Query Match      70.0%; Score 70; DB 10; Length 970;
Best Local Similarity 63.2%; Pred. No. 0.0015;
Matches 12; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXKXWQVSLQ 19
Db 433 IAGGEEACPHCWQVGLR 451

RESULT 10
US-09-925-297-871
; Sequence 871, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 871
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (57)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (70)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (87)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (106)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (112)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-871

Query Match      69.0%; Score 69; DB 10; Length 113;
Best Local Similarity 63.2%; Pred. No. 0.00023;
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXKXWQVSLQ 19
Db 29 MLCGEEARPNSWQVSLQ 47

RESULT 11
US-09-988-975A-1
; Sequence 1, Application US/09988975A
; Patent No. US20020119531A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti G.
; TITLE OF INVENTION: PROSTATE-ASSOCIATED PROTEASE ANTIBODY
; FILE REFERENCE: PF-0227-2 CIP
; CURRENT APPLICATION NUMBER: US/09/988,975A
; CURRENT FILING DATE: 2001-11-19
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PERL Program
; SEQ ID NO 1
```

```
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020119531A1 556016
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 235
; OTHER INFORMATION: unknown or other
US-09-988-975A-1

Query Match      66.0%; Score 66; DB 10; Length 283;
Best Local Similarity 66.7%; Pred. No. 0.0018;
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXKXWQVSL 18
Db 47 IVGESALFGAMPQVSL 64

RESULT 12
US-09-981-353-23
; Sequence 23, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 23
; LENGTH: 384
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 991163CD1
US-09-981-353-23

Query Match      66.0%; Score 66; DB 9; Length 384;
Best Local Similarity 66.7%; Pred. No. 0.0025;
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXKXWQVSL 18
Db 148 IVGESALFGAMPQVSL 165

RESULT 13
US-10-012-896-934
; Sequence 934, Application US/10012896
; Publication No. US20020183251A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
```



```

; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Mantanabe, Yoshihiro
; APPLICANT: Meagner, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012,896
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 934
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-896-934

```

```

Query Match          66.0%; Score 66; DB 9; Length 393;
Best Local Similarity 66.7%; Pred. No. 0.0025; 5; Indels 0; Gaps 0;
Matches 12; Conservative 1; Mismatches 5;

```

```

QY 1 IVNGEEAVPHXWQVSL 18
|| || | | | | | |
Db 157 IVGGESALPGAMPQVSL 174

```

```

RESULT 14
US-09-895-793-934
; Sequence 934, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolck, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.534C27
; CURRENT APPLICATION NUMBER: US/09/895,793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 934
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-793-934

```

```

Query Match          66.0%; Score 66; DB 9; Length 393;
Best Local Similarity 66.7%; Pred. No. 0.0025;
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 IVNGEEAVPHXWQVSL 18
|| || | | | | | |
Db 157 IVGGESALPGAMPQVSL 174

RESULT 15
US-09-895-814-934
; Sequence 934, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolck, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 934
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-814-934

Query Match          66.0%; Score 66; DB 9; Length 393;
Best Local Similarity 66.7%; Pred. No. 0.0025;
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 IVNGEEAVPHXWQVSL 18
|| || | | | | | |
Db 157 IVGGESALPGAMPQVSL 174

Search completed: February 12, 2003, 10:31:05
Job time : 3.04478 secs

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:03:55 ; Search time 8.73134 Seconds  
(without alignments)  
198.395 Million cell updates/sec

Title: US-10-036-371-5

Perfect score: 68

Sequence: 1 CGVPAIQVLSGL 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A: Geneseq 101002.\*

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	68	100.0	13	20	AAV50209	Neutrophil-activat
2	68	100.0	13	22	AAB31577	Amino terminal seq
3	68	100.0	15	20	AAV50207	Neutrophil-activat
4	68	100.0	245	21	AAV99596	Bovine chymotrypsin
5	63	92.6	146	21	AAB54191	Human pancreatic c
6	63	92.6	263	23	AAU2738	Amino acid sequenc
7	62	91.2	15	20	AAV50208	Neutrophil-activat
8	62	91.2	15	20	AAV50210	Neutrophil-activat
9	49	72.1	192	21	AAB54077	Human pancreatic c
10	49	72.1	264	21	AAB11710	Human serine prote

11	48	70.6	264	21	AAB11711	Mouse serine prote
12	46	67.6	13	20	AAV50212	Neutrophil-activat
13	42	61.8	11	20	AAV50211	Neutrophil-activat
14	42	61.8	107	22	ABG14637	Novel human diagno
15	42	61.8	107	22	AAV68565	Human novel cytol
16	42	61.8	144	22	AAU14369	Human novel protei
17	42	61.8	170	22	AAG71878	Human olfactory re
18	42	61.8	170	22	AAG72498	Human OR-like poly
19	42	61.8	327	20	AAV28279	Human G-protein co
20	42	61.8	334	22	AAG71450	Human GPCR GR1R-1
21	42	61.8	339	22	ABB12128	Human 5' EST relat
22	41	60.3	44	21	AAV64742	Amino terminal seq
23	40	58.8	13	22	AAB31575	Propionibacterium
24	40	58.8	575	22	AAU55421	S. pneumoniae deri
25	39	58.1	397	13	AAV85953	ALP of lysobacter
26	39	57.4	78	22	AAU27199	Human cardiovascular
27	39	57.4	91	22	ABG15319	Novel human diagno
28	39	57.4	168	23	AAU80535	Human G protein-co
29	39	57.4	308	22	AAG72035	Human olfactory re
30	39	57.4	319	22	AAG71816	Human olfactory re
31	39	57.4	320	22	AAG71875	Human olfactory re
32	39	57.4	320	22	AAG71935	Human olfactory re
33	39	57.4	320	22	AAG72204	Human olfactory re
34	39	57.4	323	22	AAG71442	Human olfactory re
35	39	57.4	323	22	AAG71728	Human olfactory re
36	39	57.4	327	22	AAG71444	Human olfactory re
37	39	57.4	328	22	AAG71509	Human olfactory re
38	39	57.4	330	22	AAG71933	Human olfactory re
39	39	57.4	333	22	AAG72247	Human olfactory re
40	39	57.4	334	22	AAG71421	Human olfactory re
41	39	57.4	337	22	ABG14643	Novel human diagno
42	39	57.4	337	22	AAG71423	Human olfactory re
43	39	57.4	525	15	AAV51272	Antigenic/function
44	39	57.4	609	22	ABG03578	Novel human diagno
45	39	57.4				

## ALIGNMENTS

RESULT 1

AAV50209

ID AAV50209 standard; Peptide; 13 AA.

XX AC AAV50209;

XX DT 12-JAN-2000 (first entry)

XX DE Neutrophil-activating pancreatic derived peptide 9.

XX KW Cell activation; pancreas; treatment; cardiovascular disease; trauma;

XX KW inflammatory disease; autoimmune diseases; arthritis; diabetes; stroke;

XX KW organ rejection; ischemia; Alzheimer's disease; myocardial infarction;

XX KW haemorrhagic shock; diabetic retinopathy; venous insufficiency; angina;

XX KW trauma; protease inhibitor; hypertension; sepsis.

XX OS Bos taurus.

XX PN WO9946367-A2.

XX PP 16-SEP-1999.

XX PF 11-MAR-1999; 99WO-US05247.

XX PR 11-MAR-1998; 98US-0038894.

XX PA (CELL-) CELL ACTIVATION INC.

XX PA (REGC ) UNIV CALIFORNIA.

XX PA (SCRI ) SCRIPPS RES INST.

XX PI Stoughton RB, Schmid-Schonbein GW, Hugli TE, Kistler E;

XX WPI; 1999-580234/49.

```

XX PT Use of cell activating compositions in developing products for
XX PT diagnosis and treatment of e.g. cardiovascular, inflammatory,
XX PT autoimmune or Alzheimer's disease, trauma, arthritis, organ rejection,
XX PT diabetes, stroke or ischemia -
XX PS Example 9; Page 182; 184pp; English.
XX CC This invention describes a novel method for the use and preparation of
XX CC cell activating compositions which involves preparing a cell activating
XX CC composition comprising (a) homogenizing pancreatic tissue in buffer at
XX CC about neutral or higher pH to produce a homogenate; (b) removing the
XX CC particulates from the homogenate; (c) optionally incubating the
XX CC resulting homogenate, with particulates removed, with a protease; and
XX CC (d) fractionating the homogenate and selecting fractions that exhibit
XX CC cell activation activity. The methods can be used for improving
XX CC treatment outcome or reducing risk of treatment of e.g. cardiovascular
XX CC disease, inflammatory disease, trauma, autoimmune diseases, arthritis,
XX CC organ rejection, diabetes and diabetic complications, stroke, ischemia,
XX CC Alzheimer's disease, myocardial infarction, hemorrhagic shock, diabetic
XX CC retinopathy, diabetes, venous insufficiency, unstable angina or trauma.
XX CC They can be used in the veterinary treatment of a non-human subject.
XX CC Protease inhibitors can be used to lower cell activation resulting from
XX CC these diseases and deficiencies. The detection of an elevated level of
XX CC hydrogen peroxide can be used to detect an inflammatory condition. An
XX CC elevated level of hydrogen peroxide in plasma or whole blood and in the
XX CC presence of superoxide dismutase (SOD) indicates leukocyte up
XX CC regulation, e.g. indicative of the onset of an acute cardiovascular
XX CC disorders, such as disease onset or ischemic complications. An elevated
XX CC level of hydrogen peroxide in plasma or whole blood and a low level in
XX CC the presence of SOD is indicative of a chronic or immune compromised
XX CC condition e.g. hypertension or sepsis. AAY50201-Y50334 represent peptides
XX CC used in the method of the invention.
XX SQ Sequence 13 AA;
    Query Match 100.0%; Score 68; DB 20; Length 13;
    Best Local Similarity 100.0%; Pred. No. 5.1e-05;
    Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGVPAIQPVLSGL 13
Db 1 CGVPAIQPVLSGL 13
    |||||
RESULT 2
AAB31577
ID AAB31577 standard; peptide; 13 AA.
AC AAB31577;
XX 20-APR-2001 (first entry)
XX Amino terminal sequence of bovine trypsin A chain.
XX Fish; serine proteinase; pain; acute inflammation; chronic inflammation;
XX arthritis; inflamed joint; bursitis; osteoarthritis; septic arthritis;
XX rheumatoid arthritis; juvenile rheumatoid arthritis; fibromyalgia;
XX systemic lupus erythematosus; phlebitis; tendinitis; rash; psoriasis;
XX acne; eczema; facial seborrheic eczema; foreskin infection;
XX athlete's foot; fistulae infection; ulcer; navel infection; wrinkle;
XX scar; keloid; boil; wart; allergic itch; hemorrhoid; wound;
XX fungal infection; autoimmune disease.
XX Bos sp.
XX WO200078332-A2.
XX 28-DEC-2000.
XX 15-JUN-2000; 2000WO-1500005.
XX 18-JUN-1999; 99IS-0005086.

```

```

XX PA (BUAR/) BJARNASON J B.
XX PJ Bjarnason JB;
XX DR WPI; 2001-091493/10.
XX PT Fish serine proteinase, useful as a cosmetic, medicament for treating
XX PT eczema, psoriasis, arthritis, and in the manufacture of the medicament
XX PT for treating preventing pathogenic diseases involving receptor
XX PT mediated binding -
XX PS Disclosure; Page 5; 38pp; English.
XX CC The specification describes a fish serine proteinase. The proteinases
XX CC are useful as medicaments, for treating and preventing a disease in a
XX CC human or an animal such as pain, acute inflammation, chronic
XX CC inflammation, arthritis, inflamed joints, bursitis, osteoarthritis,
XX CC rheumatoid arthritis, juvenile rheumatoid arthritis, septic arthritis,
XX CC fibromyalgia, systemic lupus erythematosus, phlebitis, tendinitis, rash,
XX CC psoriasis, acne, eczema, facial seborrheic eczema, eczema of the hands,
XX CC face or neck, foreskin infections, athlete's foot, fistulae infections,
XX CC infected topical ulcers, navel infections in newborns, wrinkles, scars,
XX CC keloids, boils, warts and allergic itch, hemorrhoids, wounds, wound
XX CC infections, wounds from burns, fungal infections and immunological and
XX CC autoimmune diseases. They are also useful for removing dead or peeling
XX CC skin from otherwise healthy skin, and for treating or preventing a
XX CC disease in which pathogenesis is caused by bacteria, virus, fungus,
XX CC parasite or a protozoan or a receptor mediated binding is involved.
XX CC The present sequence represents the amino terminal of bovine trypsin A
XX CC chain.
XX SQ Sequence 13 AA;
    Query Match 100.0%; Score 68; DB 22; Length 13;
    Best Local Similarity 100.0%; Pred. No. 5.1e-05;
    Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGVPAIQPVLSGL 13
Db 1 CGVPAIQPVLSGL 13
    |||||
RESULT 3
AAY50207
ID AAY50207 standard; Peptide; 15 AA.
XX AAY50207;
XX 12-JAN-2000 (first entry)
XX Neutrophil-activating pancreatic derived peptide 7.
XX Cell activation; pancreas; treatment; cardiovascular disease; trauma;
XX inflammatory disease; autoimmune diseases; arthritis; diabetes; stroke;
XX organ rejection; ischemia; Alzheimer's disease; myocardial infarction;
XX hemorrhagic shock; diabetic retinopathy; venous insufficiency; angina;
XX trauma; protease inhibitor; hypertension; sepsis.
XX Bos taurus.
XX WO9946367-A2.
XX 16-SEP-1999.
XX 11-MAR-1999; 99WO-US05247.
XX 11-MAR-1998; 98US-0038894.
XX (CELL-) CELL ACTIVATION INC.
XX (REGC) UNIV CALIFORNIA.
XX (SCRI) SCRIPPS RES INST.

```

PI Stoughton RB, Schmid-Schonbein GW, Hugli TE, Kistler E;  
 XX WPI; 1999-580234/49.  
 XX  
 XX Use of cell activating compositions in developing products for  
 PT diagnosis and treatment of e.g. cardiovascular, inflammatory,  
 PT autoimmune or Alzheimer's disease, trauma, arthritis, organ rejection,  
 PT diabetes, stroke or ischemia -  
 XX  
 PS Example 9; Page 182; 184pp; English.  
 XX  
 XX This invention describes a novel method for the use and preparation of  
 CC cell activating compositions which involves preparing a cell activating  
 CC composition comprising (a) homogenizing pancreatic tissue in buffer at  
 CC about neutral or higher pH to produce a homogenate; (b) removing  
 CC particulates from the homogenate; (c) optionally incubating the  
 CC resulting homogenate, with particulates removed, with a protease; and  
 CC (d) fractionating the homogenate and selecting fractions that exhibit  
 CC cell activation activity. The methods can be used for improving  
 CC treatment outcome or reducing risk of treatment of e.g. cardiovascular  
 CC disease, inflammatory disease, trauma, autoimmune diseases, arthritis,  
 CC organ rejection, diabetes and diabetic complications, stroke, ischemia,  
 CC Alzheimer's disease, myocardial infarction, haemorrhagic shock, diabetic  
 CC retinopathy, diabetes, venous insufficiency, unstable angina or trauma.  
 CC They can be used in the veterinary treatment of a non-human subject.  
 CC Protease inhibitors can be used to lower cell activation resulting from  
 CC these diseases and deficiencies. The detection of an elevated level of  
 CC hydrogen peroxide can be used to detect an inflammatory condition. An  
 CC elevated level of hydrogen peroxide in plasma or whole blood and in the  
 CC presence of superoxide dismutase (SOD) indicates leukocyte up  
 CC regulation, e.g. indicative of the onset of an acute cardiovascular  
 CC disorders, such as disease onset or ischemic complications. An elevated  
 CC level of hydrogen peroxide in plasma or whole blood and a low level in  
 CC the presence of SOD is indicative of a chronic or immune compromised  
 CC condition e.g. hypertension or sepsis. AAY50201-Y50334 represent peptides  
 CC used in the method of the invention.  
 XX  
 XX  
 SQ Sequence 15 AA;  
 Query Match 100.0%; Score 68; DB 20; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 6e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 CGVPAIQPVLSGL 13  
 Db | | | | | | | | | | | | | | | |  
 1 CGVPAIQPVLSGL 13  
 RESULT 4  
 AAY99596  
 ID AAY99596 standard; protein; 245 AA.  
 AC AAY99596;  
 XX  
 XX  
 DT 13-SEP-2000 (first entry)  
 XX  
 DE Bovine chymotrypsinogen A.  
 XX  
 KW Bovine; plasminogen activator; cardiant; thrombolytic;  
 KW heart attack; stroke; blood clotting disorder.  
 XX  
 OS Bos taurus.  
 XX  
 PN WO200032759-A1.  
 XX  
 PD 08-JUN-2000.  
 XX  
 PF 06-MAY-1999; 99WO-0509991.  
 XX  
 PR 02-DEC-1998; 98US-0110588.  
 XX  
 PI (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 PA  
 XX

PI Lin X, Zhang XC, Tang JTN;  
 XX WPI; 2000-422975/36.  
 XX  
 XX Polypeptide with plasminogen activator activity useful as thrombolytic  
 PT agent for treating blood clot disorders e.g. heart attack, comprises 10  
 PT amino acid peptide fragment for recognition or activation of  
 PT plasminogen -  
 XX  
 PS Disclosure; Page 35-36; 41pp; English.  
 XX  
 XX The present sequence is bovine chymotrypsinogen. It was included in  
 CC a review of sequence homologies of several plasminogen activators.  
 CC Plasminogen is the principal serine protease zymogen in the  
 CC extracellular fluids of vertebrates. Its active form, plasmin, is  
 CC implicated in pericellular proteolysis associated with a wide range of  
 CC physiological and pathological processes. Plasminogen expression is  
 CC regulated by plasminogen activators which hydrolyse a peptide bond in  
 CC plasminogen to convert it to plasmin or form tight binding complexes  
 CC with plasminogen to spontaneously convert it to plasmin. The sequence  
 CC homology analysis has identified a six amino acid peptide involved in  
 CC plasminogen activation. This peptide is particularly useful when inserted  
 CC between amino acid residues 644 and 645 of full length human  
 CC plasminogen. Novel plasminogen activators have been made based upon the  
 CC plasminogen activation/recognition site of plasminogen binding proteins.  
 CC The polypeptides are useful in preparing thrombolytic agents for treating  
 CC blood clotting disorders such as heart attack.  
 XX  
 SQ Sequence 245 AA;  
 Query Match 100.0%; Score 68; DB 21; Length 245;  
 Best Local Similarity 100.0%; Pred. No. 0.0012;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 CGVPAIQPVLSGL 13  
 Db | | | | | | | | | | | | | | | |  
 1 CGVPAIQPVLSGL 13  
 RESULT 5  
 AAB54191  
 ID AAB54191 standard; Protein; 146 AA.  
 XX  
 AC AAB54191;  
 XX  
 XX  
 DT 09-MAR-2001 (first entry)  
 XX  
 DE Human pancreatic cancer antigen protein sequence SEQ ID NO:643.  
 XX  
 KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;  
 KW detection; diagnosis; identification; cytostatic; neuroprotective;  
 KW nontropic; immunomodulatory; relaxant; contraceptive; gynaecological;  
 KW antiinflammatory; cardiant; gene therapy; chromosome mapping;  
 KW linkage analysis; tissue identification; tissue typing; forensic;  
 KW neural; immune system; muscular; reproductive; gastrointestinal;  
 KW pulmonary; cardiovascular; renal; proliferative.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2000055320-A1.  
 XX  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 08-MAR-2000; 2000WO-US05989.  
 XX  
 PR 12-MAR-1999; 99US-0124270.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 XX WPI; 2000-579444/54.  
 DR  
 DR N-PSDB; AAC98956.

XX New nucleic acid that is a pancreatic cancer antigen for preventing,  
PT treating, or ameliorating a medical condition, particular pancreatic  
PT cancer, or for use in assays for diagnosing a pathological condition -  
XX  
PS Claim 11; Page 1081; 1379pp; English.  
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated  
CC proteins, called pancreatic cancer antigens, given in AAB54008 to  
CC AAB54466. The human pancreatic cancer antigens have cytostatic,  
CC neuroprotective, nontropic, immunomodulatory, relaxant, contraceptive,  
CC gynaecological, cardiant and antiinflammatory activities, and can be used  
CC in gene therapy. The polynucleotide and proteins can be used for  
CC preventing, treating, or ameliorating a medical condition or in assays  
CC for diagnosing a pathological condition or a susceptibility to one in a  
CC subject. Binding partners to the proteins and the activity of the  
CC proteins can be identified. The pancreatic cancer antigens can be used to  
CC detect, treat or prevent pancreatic disorders, especially cancer.  
CC Agonists and antagonists to the antigens can be screened for. The  
CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
CC acid hybridisation probes that can be used in chromosome mapping, linkage  
CC analysis, tissue identification and/or typing and a variety of forensic  
CC and diagnostic methods. The proteins can be used to generate antibodies  
CC which are used to purify, detect and target the polypeptides, including  
CC both in vivo and in vitro diagnostic and therapeutic methods. The  
CC proteins can be used to treat or prevent neural, immune system, muscular,  
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
CC sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 146 AA;

Query Match 92.6%; Score 63; DB 21; Length 146;  
Best Local Similarity 92.3%; Pred. No. 0.0051;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 CGVPAIQPVLSGL 13  
Db 23 CGVPAIHPVLSGL 35

RESULT 6  
AAU82738  
ID AAU82738 standard; Protein; 263 AA.  
XX  
AC AAU82738;  
XX  
DT 23-APR-2002 (first entry)  
XX  
DE Amino acid sequence of novel human protease #37.  
XX Human; protease; cancer; immune-related disorder; cardiovascular disease;  
XX neuronal-associated disease; metabolic disorder; inflammatory disorder;  
XX nervous system disorder; sexual dysfunction; pain; mood disorder;  
XX hypertension; psychotic disorder; neurological disorder; dyskinesia;  
XX viral infection; human immunodeficiency virus; HIV; non-viral infection;  
XX ocular disease; cytostatic; enzyme.  
XX  
OS Homo sapiens.  
XX  
PN WO200200860-A2.  
XX  
XX 03-JAN-2002.  
XX  
PD 26-JUN-2001; 2001WO-US20171.  
XX  
XX 26-JUN-2000; 2000US-214047P.  
XX  
XX (SUGB-) SUGEN INC.  
XX  
XX Plowman G, Whyte D, Sudarshanam S, Manning G, Caenepeel S;  
XX Charyczak G;  
XX

DR WPI; 2002-139913/18.  
DR N-PSDB; ABK31780.  
PT Nucleic acids encoding novel human proteases, useful for useful for  
PT treating diseases and disorders such as cancers, immune-related  
PT diseases and disorders, cardiovascular disease (e.g. restenosis) and  
PT inflammatory disorders -  
XX  
PS Claim 6; Fig 2M; 313pp; English.  
XX  
XX The present invention relates to the isolation of novel human  
XX proteases, and the nucleic acids encoding them. The sequences of  
CC the invention are useful for treating diseases and disorders such as  
CC cancers (e.g. breast, colon, lung), immune-related diseases and disorders  
CC (e.g. inflammatory diseases and asthma), cardiovascular diseases  
CC (e.g. restenosis and coronary thrombosis), brain or neuronal-associated  
CC diseases, metabolic disorders (e.g. diabetes, obesity), inflammatory  
CC disorders (e.g. rheumatoid arthritis and psoriasis), central or  
CC peripheral nervous system diseases, migraines, pain, sexual dysfunction,  
CC mood disorders, attention disorders, cognitive disorders, hypotension,  
CC hypertension, psychotic disorders, neurological disorders  
CC (e.g. Alzheimer's disease, Parkinson's disease) and dyskinesias.  
CC The nucleic acids and polypeptides are also useful for treating viral  
CC infections caused by human immunodeficiency virus (HIV), and non-viral  
CC infections such as ocular disease (e.g. glaucoma) and macular  
CC degeneration. AAU82702-AAU82760 represent the novel human proteases of  
CC the invention.  
XX  
SQ Sequence 263 AA;

Query Match 92.6%; Score 63; DB 23; Length 263;  
Best Local Similarity 92.3%; Pred. No. 0.0096;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 CGVPAIQPVLSGL 13  
Db 19 CGVPAIHPVLSGL 31

RESULT 7  
AAU50208  
ID AAU50208 standard; Peptide; 15 AA.  
XX  
AC AAU50208;  
XX  
DT 12-JAN-2000 (first entry)  
XX  
DE Neutrophil-activating pancreatic derived peptide 8.  
XX  
XX Cell activation; pancreas; treatment; cardiovascular disease; trauma;  
XX inflammatory disease; autoimmune diseases; arthritis; diabetes; stroke;  
XX organ rejection; ischemia; Alzheimer's disease; myocardial infarction;  
XX haemorrhagic shock; diabetic retinopathy; venous insufficiency; angina;  
XX trauma; protease inhibitor; hypertension; sepsis.  
XX  
OS Unidentified.  
XX  
PN WO9946367-A2.  
XX  
XX 16-SEP-1999.  
XX  
XX 11-MAR-1999; 99WO-US05247.  
XX  
XX 11-MAR-1998; 98US-0038894.  
XX  
XX (CELL-) CELL ACTIVATION INC.  
XX (REGC ) UNIV CALIFORNIA.  
XX (SCRI ) SCRIPPS RES INST.  
XX  
XX Stoughton RB, Schmid-Schonbein GW, Hugli TE, Kistler E;  
XX WPI, 1999-580234/49.  
XX

PT Use of cell activating compositions in developing products for  
 PI diagnosis and treatment of e.g. cardiovascular, inflammatory,  
 PT autoimmune or Alzheimer's disease, trauma, arthritis, organ rejection,  
 XX diabetes, stroke or ischemia -  
 PS Example 9; Page 182; 184pp; English.

XX This invention describes a novel method for the use and preparation of  
 CC cell activating compositions which involves preparing a cell activating  
 CC composition comprising (a) homogenizing pancreatic tissue in buffer at  
 CC about neutral or higher pH to produce a homogenate; (b) removing  
 CC particulates from the homogenate; (c) optionally incubating the  
 CC resulting homogenate, with particulates removed, with a protease; and  
 CC (d) fractionating the homogenate and selecting fractions that exhibit  
 CC cell activation activity. The methods can be used for improving  
 CC treatment outcome or reducing risk of treatment of e.g. cardiovascular  
 CC disease, inflammatory disease, trauma, autoimmune diseases, arthritis,  
 CC organ rejection, diabetes and diabetic complications, stroke, ischemia,  
 CC Alzheimer's disease, myocardial infarction, haemorrhagic shock, diabetic  
 CC retinopathy, diabetes, venous insufficiency, unstable angina or trauma.  
 CC They can be used in the veterinary treatment of a non-human subject.  
 CC Protease inhibitors can be used to lower cell activation resulting from  
 CC these diseases and deficiencies. The detection of an elevated level of  
 CC hydrogen peroxide can be used to detect an inflammatory condition. An  
 CC elevated level of hydrogen peroxide in plasma or whole blood and in the  
 CC presence of superoxide dismutase (SOD) indicates leukocyte up  
 CC regulation, e.g. indicative of the onset of an acute cardiovascular  
 CC disorders, such as disease onset or ischemic complications. An elevated  
 CC level of hydrogen peroxide in plasma or whole blood and a low level in  
 CC the presence of SOD is indicative of a chronic or immune compromised  
 CC condition e.g. hypertension or sepsis. AAY50201-Y50334 represent peptides  
 CC used in the method of the invention.

XX Sequence 15 AA;

Query Match 91.2%; Score 62; DB 20; Length 15;  
 Best Local Similarity 92.3%; Pred. No. 0.00064;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLSGL 13  
 DQ 1 CGVPAIPPVLSGL 13

RESULT 8  
 AAY50210  
 ID AAY50210 standard; Peptide; 15 AA.

AC AAY50210;

DT 12-JAN-2000 (first entry)

DE Neutrophil-activating pancreatic derived peptide 10.

KW Cell activation; pancreas; treatment; cardiovascular disease; trauma;  
 KW inflammatory disease; autoimmune diseases; arthritis; diabetes; stroke;  
 KW organ rejection; ischemia; Alzheimer's disease; myocardial infarction;  
 KW haemorrhagic shock; diabetic retinopathy; venous insufficiency; angina;  
 KW trauma; protease inhibitor; hypertension; sepsis.

OS Unidentified.

FN WO9946367-A2.

PD 16-SEP-1999.

PF 11-MAR-1999; 99MO-US05247.

PR 11-MAR-1998; 98US-0038894.

XX (CELL-) CELL ACTIVATION INC.

PA (REGC) UNIV CALIFORNIA.

PA (SCRI) SCRIPPS RES INST.

XX Stoughton RB, Schmid-Schonbein GW, Hugli TE, Kistler E;  
 XX WPT; 1999-580234/49.  
 XX Use of cell activating compositions in developing products for  
 PT diagnosis and treatment of e.g. cardiovascular, inflammatory,  
 PT autoimmune or Alzheimer's disease, trauma, arthritis, organ rejection,  
 XX diabetes, stroke or ischemia -  
 PS Example 9; Page 182; 184pp; English.

XX This invention describes a novel method for the use and preparation of  
 CC cell activating compositions which involves preparing a cell activating  
 CC composition comprising (a) homogenizing pancreatic tissue in buffer at  
 CC about neutral or higher pH to produce a homogenate; (b) removing  
 CC particulates from the homogenate; (c) optionally incubating the  
 CC resulting homogenate, with particulates removed, with a protease; and  
 CC (d) fractionating the homogenate and selecting fractions that exhibit  
 CC cell activation activity. The methods can be used for improving  
 CC treatment outcome or reducing risk of treatment of e.g. cardiovascular  
 CC disease, inflammatory disease, trauma, autoimmune diseases, arthritis,  
 CC organ rejection, diabetes and diabetic complications, stroke, ischemia,  
 CC Alzheimer's disease, myocardial infarction, haemorrhagic shock, diabetic  
 CC retinopathy, diabetes, venous insufficiency, unstable angina or trauma.  
 CC They can be used in the veterinary treatment of a non-human subject.  
 CC Protease inhibitors can be used to lower cell activation resulting from  
 CC these diseases and deficiencies. The detection of an elevated level of  
 CC hydrogen peroxide can be used to detect an inflammatory condition. An  
 CC elevated level of hydrogen peroxide in plasma or whole blood and in the  
 CC presence of superoxide dismutase (SOD) indicates leukocyte up  
 CC regulation, e.g. indicative of the onset of an acute cardiovascular  
 CC disorders, such as disease onset or ischemic complications. An elevated  
 CC level of hydrogen peroxide in plasma or whole blood and a low level in  
 CC the presence of SOD is indicative of a chronic or immune compromised  
 CC condition e.g. hypertension or sepsis. AAY50201-Y50334 represent peptides  
 CC used in the method of the invention.

XX Sequence 15 AA;

Query Match 91.2%; Score 62; DB 20; Length 15;  
 Best Local Similarity 92.3%; Pred. No. 0.00064;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLSGL 13  
 DQ 1 CGVPAIPPVLSGL 13

RESULT 9  
 AAB54077

ID AAB54077 standard; Protein; 192 AA.

AC AAB54077;

DT 09-MAR-2001 (first entry)

DE Human pancreatic cancer antigen protein sequence SEQ ID NO:529.

KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;  
 KW detection; diagnosis; identification; cytostatic; neuroprotective;  
 KW neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;  
 KW antiinflammatory; cardiant; gene therapy; chromosome mapping;  
 KW linkage analysis; tissue identification; tissue typing; forensic;  
 KW neural; immune system; muscular; reproductive; gastrointestinal;  
 KW pulmonary; cardiovascular; renal; proliferative.

OS Homo sapiens.

PN WO200055320-A1.

XX 21-SEP-2000.

Wed Feb 12 11:59:39 2003

PF 08-MAR-2000; 2000WO-US05989.  
XX 12-MAR-1999; 99US-0124270.  
PR (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Ruben SM;  
PI WPI; 2000-579444/54.  
XX N-PSDB; AAC98842.  
DR New nucleic acid that is a pancreatic cancer antigen for preventing,  
XX treating, or ameliorating a medical condition, particular pancreatic  
PT cancer, or for use in assays for diagnosing a pathological condition -  
PT  
XX  
XX Claim 11; Page 966; 1379pp; English.  
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated  
CC proteins, called pancreatic cancer antigens, given in AAS4008 to  
CC AAB54466. The human pancreatic cancer antigens have cytostatic,  
CC neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,  
CC gynaecological, cardiant and antiinflammatory activities, and can be used  
CC in gene therapy. The polynucleotide and proteins can be used for  
CC preventing, treating, or ameliorating a medical condition or in assays  
CC for diagnosing a pathological condition or a susceptibility to one in a  
CC subject. Binding partners to the proteins and the activity of the  
CC proteins can be identified. The pancreatic cancer antigens can be used to  
CC detect, treat or prevent pancreatic disorders, especially cancer.  
CC Agonists and antagonists to the antigens can be screened for. The  
CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
CC acid hybridisation probes that can be used in chromosome mapping, linkage  
CC analysis, tissue identification and/or typing and a variety of forensic  
CC and diagnostic methods. The proteins can be used to generate antibodies  
CC which are used to purify, detect and target the polypeptides, including  
CC both in vivo and in vitro diagnostic and therapeutic methods. The  
CC proteins can be used to treat or prevent neural, immune system, muscular,  
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
CC sequences used in the exemplification of the present invention.  
XX  
XX Sequence 192 AA;  
SQ  
Query Match 72.1%; Score 49; DB 21; Length 192;  
Best Local Similarity 72.7%; Pred. No. 1.7;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CGVPAIQPVLS 11  
DB 15 CGIPAIKPALS 25  
RESULT 10  
AAB11710  
ID AAB11710 standard; Protein; 264 AA.  
XX  
AC AAB11710;  
XX  
DT 23-OCT-2000 (first entry)  
XX  
DE Human serine protease BSSP5 (hBSSP5) SEQ ID NO:2.  
XX  
KW BSSP5; serine protease; human; hBSSP5; mouse; mBSSP5; brain;  
KW diagnostic marker; antibody; transgenic animal; Alzheimer's disease;  
KW epilepsy; cancer; inflammation; infertility; pancreatitis;  
KW prostatic hypertrophy.  
XX  
OS Homo sapiens.  
XX  
PN WO200031243-A1.  
XX  
PD 02-JUN-2000.  
XX  
PF 19-NOV-1999; 99WO-JP06473.

XX 20-NOV-1998; 98JP-0347806.  
XX (FUSO) FUSO PHARM IND LTD.  
XX  
XX Uemura H, Okui A, Kominami K, Yamaguchi N, Miteui S;  
XX WPI; 2000-400058/34.  
XX N-PSDB; AAB61733.  
DR  
XX Serine proteases BSSP5, useful in detecting homologs, mutants and  
XX polymorphic variants as markers for diagnosis of e.g. Alzheimer's  
PT disease, epilepsy, cancer and inflammation, using blood, urine,  
PT pancreas or other tissues -  
XX  
XX Claim 1; Page 51-52; 70pp; Japanese.  
XX  
XX The invention relates to novel serine proteases designated BSSP5  
CC (AAB11710-B11711), and to nucleic acids encoding them (AAB61733-A61734).  
CC The invention also relates to vectors and transformants comprising BSSP5  
CC nucleic acids; transgenic animals in which the expression level of BSSP5  
CC can be varied; and an mBSSP5 knockout mouse. The invention additionally  
CC encompasses anti-BSSP5 antibodies and methods of production of such  
CC antibodies, methods of BSSP5 detection using the antibodies, and the  
CC use of BSSP5 proteins or fragments as diagnostic markers for certain  
CC medical conditions, e.g., pancreatitis. A method for detecting  
CC pancreaticitis comprising measuring BSSP5 concentration in the blood or  
CC urine, and a pancreaticitis diagnostic agent containing an anti-BSSP5  
CC antibody is also disclosed. Nucleotides encoding BSSP5 were initially  
CC isolated in a human brain cDNA library using degenerate PCR primers  
CC (AAB61744-A61745) based on conserved regions of serine proteases. The  
CC BSSP5 serine proteases, mutants and polymorphic variants in biological  
CC samples (e.g., blood, urine, brain, prostate gland, placenta, testis,  
CC pancreas and spleen) as diagnostic markers for conditions such as  
CC Alzheimer's disease, epilepsy, cancer, inflammation, infertility,  
CC pancreatitis and prostatic hypertrophy. Sequence AAB11710 represents  
CC human BSSP5 (hBSSP5), and sequence AAB11711 represents murine BSSP5  
CC (mBSSP5).  
XX  
XX Sequence 264 AA;  
SQ  
Query Match 72.1%; Score 49; DB 21; Length 264;  
Best Local Similarity 72.7%; Pred. No. 2.4;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CGVPAIQPVLS 11  
DB 19 CGIPAIKPALS 29  
RESULT 11  
AAB11711  
ID AAB11711 standard; Protein; 264 AA.  
XX  
AC AAB11711;  
XX  
DT 23-OCT-2000 (first entry)  
XX  
DE Mouse serine protease BSSP5 (mBSSP5) SEQ ID NO:4.  
XX  
KW BSSP5; serine protease; human; hBSSP5; mouse; mBSSP5; brain;  
KW diagnostic marker; antibody; transgenic animal; Alzheimer's disease;  
KW epilepsy; cancer; inflammation; infertility; pancreatitis;  
KW prostatic hypertrophy.  
XX  
OS Mus sp.  
XX  
PN WO200031243-A1.  
XX  
PD 02-JUN-2000.  
XX  
PF 19-NOV-1999; 99WO-JP06473.



XX PR 20-NOV-1998; 98JP-0347806.  
 XX PA (FUSO ) FUSO PHARM IND LTD.  
 XX PI Uemura H, Okui A, Kominami K, Yamaguchi N, Mitsui S;  
 XX DR WPI: 2000-40058/34.  
 XX DR N-PSDB; AAA61734.  
 XX PT Serine proteases BSSP5, useful in detecting homologs, mutants and  
 PT polymorphic variants as markers for diagnosis of e.g. Alzheimer's  
 PT disease, epilepsy, cancer and inflammation, using blood, urine,  
 PT pancreas or other tissues -  
 XX PS Claim 3; Page 55-56; 70pp; Japanese.  
 XX CC The invention relates to novel serine proteases designated BSSP5  
 CC (AAB11710-B11711), and to nucleic acids encoding them (AAA61733-A61734).  
 CC The invention also relates to vectors and transformants comprising BSSP5  
 CC nucleic acids; transgenic animals in which the expression level of BSSP5  
 CC can be varied; and an mBSSP5 knockout mouse. The invention additionally  
 CC encompasses anti-BSSP5 antibodies and methods of production of such  
 CC antibodies, methods of BSSP5 detection using the antibodies, and the  
 CC use of BSSP5 proteins or fragments as diagnostic markers for certain  
 CC medical conditions, e.g., pancreatitis. A method for detecting  
 CC pancreatitis comprising measuring BSSP5 concentration in the blood or  
 CC urine, and a pancreatitis diagnostic agent containing an anti-BSSP5  
 CC antibody is also disclosed. Nucleotides encoding BSSP5 were initially  
 CC isolated in a human brain cDNA library using degenerate PCR primers  
 CC (AAA61744-A61745) based on conserved regions of serine proteases. The  
 CC BSSP5 serine proteases and nucleotides encoding them are useful in  
 CC detecting homologs, mutants and polymorphic variants in biological  
 CC samples (e.g. blood, urine, brain, prostate gland, placenta, testis,  
 CC pancreas and spleen) as diagnostic markers for conditions such as  
 CC Alzheimer's disease, epilepsy, cancer, inflammation, infertility,  
 CC pancreatitis and prostatic hypertrophy. Sequence AAB11710 represents  
 CC human BSSP5 (hBSSP5), and sequence AAB11711 represents murine BSSP5  
 CC (mBSSP5).  
 XX SQ Sequence 264 AA;  
 Query Match 70.6%; Score 48; DB 21; Length 264;  
 Best Local Similarity 81.8%; Pred. No. 3.6;  
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CGVPAIQPVLS 11  
 Db ||||| ||  
 19 CGVPAITPALS 29  
 RESULT 12  
 AAY50212  
 ID AAY50212 standard; Peptide; 13 AA.  
 AC AAY50212;  
 XX 12-JAN-2000 (first entry)  
 DE Neutrophil-activating pancreatic derived peptide 12.  
 XX Cell activation; pancreas; treatment; cardiovascular disease; trauma;  
 KW inflammatory disease; autoimmune diseases; arthritis; diabetes; stroke;  
 KW organ rejection; ischemia; Alzheimer's disease; myocardial infarction;  
 KW haemorrhagic shock; diabetic retinopathy; venous insufficiency; angina;  
 KW trauma; protease inhibitor; hypertension; sepsis.  
 XX Unidentified.  
 OS  
 KW  
 XX  
 FN MO9946367-A2.  
 XX  
 PD 16-SEP-1999.  
 XX

PF 11-MAR-1999; 99WO-US05247.  
 XX PR 11-MAR-1998; 98US-0038894.  
 XX PA (CELL-) CELL ACTIVATION INC.  
 PA (REGC ) UNIV CALIFORNIA.  
 PA (SCRI ) SCRIPPS RES INST.  
 XX PI Stoughton RB, Schmid-Schonbein GW, Hugli TE, Kistler E;  
 XX WPI: 1999-580234/49.  
 DR Use of cell activating compositions in developing products for  
 PT diagnosis and treatment of e.g. cardiovascular, inflammatory,  
 PT autoimmune or Alzheimer's disease, trauma, arthritis, organ rejection,  
 PT diabetes, stroke or ischemia -  
 XX Example 9; Page 182; 184pp; English.  
 XX This invention describes a novel method for the use and preparation of  
 CC cell activating compositions which involves preparing a cell activating  
 CC composition comprising (a) homogenizing pancreatic tissue in buffer at  
 CC about neutral or higher pH to produce a homogenate; (b) removing  
 CC particulates from the homogenate; (c) optionally incubating the  
 CC resulting homogenate, with particulates removed, with a protease; and  
 CC (d) fractionating the homogenate and selecting fractions that exhibit  
 CC cell activation activity. The methods can be used for improving  
 CC treatment outcome or reducing risk of treatment of e.g. cardiovascular  
 CC disease, inflammatory disease, trauma, autoimmune diseases, arthritis,  
 CC organ rejection, diabetes and diabetic complications, stroke, ischemia,  
 CC Alzheimer's disease, myocardial infarction, haemorrhagic shock, diabetic  
 CC retinopathy, diabetes, venous insufficiency, unstable angina or trauma.  
 CC They can be used in the veterinary treatment of a non-human subject.  
 CC Protease inhibitors can be used to lower cell activation resulting from  
 CC these diseases and deficiencies. The detection of an elevated level of  
 CC hydrogen peroxide can be used to detect an inflammatory condition. An  
 CC elevated level of hydrogen peroxide in plasma or whole blood and in the  
 CC presence of superoxide dismutase (SOD) indicates leukocyte up  
 CC regulation, e.g. indicative of the onset of an acute cardiovascular  
 CC disorders, such as disease onset or ischemic complications. An elevated  
 CC level of hydrogen peroxide in plasma or whole blood and a low level in  
 CC the presence of SOD is indicative of a chronic or immune compromised  
 CC condition e.g. hypertension or sepsis. AAY50201-Y50334 represent peptides  
 CC used in the method of the invention.  
 XX SQ Sequence 13 AA;  
 Query Match 67.6%; Score 46; DB 20; Length 13;  
 Best Local Similarity 80.0%; Pred. No. 0.31;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CGVPAIQPVL 10  
 Db ||||| |  
 1 CGVPAIKPAL 10  
 RESULT 13  
 AAY50211  
 ID AAY50211 standard; Peptide; 11 AA.  
 XX AC AAY50211;  
 XX 12-JAN-2000 (first entry)  
 DE Neutrophil-activating pancreatic derived peptide 11.  
 XX Cell activation; pancreas; treatment; cardiovascular disease; trauma;  
 KW inflammatory disease; autoimmune diseases; arthritis; diabetes; stroke;  
 KW organ rejection; ischemia; Alzheimer's disease; myocardial infarction;  
 KW haemorrhagic shock; diabetic retinopathy; venous insufficiency; angina;  
 KW trauma; protease inhibitor; hypertension; sepsis.  
 XX Unidentified.  
 OS

XX OS Homo sapiens.  
XX PN WO9946367-A2.  
XX PD 16-SEP-1999.  
XX PP 11-MAR-1999; 99WO-US05247.  
XX PR 11-MAR-1998; 98US-0038894.  
XX PA (CELL-) CELL ACTIVATION INC.  
XX PA (REGC ) UNIV CALIFORNIA.  
XX PA (SCRI ) SCRIPPS RES INST.  
XX PI Stoughton RB, Schmid-Schonbein GW, Hugli TE, Kistler E;  
XX PI WPI; 1999-580234/49.  
XX DR Use of cell activating compositions in developing products for  
XX PT diagnosis and treatment of e.g. cardiovascular, inflammatory,  
XX PT autoimmune or Alzheimer's disease, trauma, arthritis, organ rejection,  
XX PT diabetes, stroke or ischemia -  
XX PS Example 9; Page 182; 184pp; English.  
XX CC This invention describes a novel method for the use and preparation of  
XX CC cell activating compositions which involves preparing a cell activating  
XX CC composition comprising (a) homogenizing pancreatic tissue in buffer at  
XX CC about neutral or higher pH to produce a homogenate; (b) removing the  
XX CC particulates from the homogenate; (c) optionally incubating the  
XX CC resulting homogenate, with particulates removed, with a protease; and  
XX CC (d) fractionating the homogenate and selecting fractions that exhibit  
XX CC cell activation activity. The methods can be used for improving  
XX CC treatment outcome or reducing risk of treatment of e.g. cardiovascular  
XX CC disease, inflammatory disease, trauma, autoimmune diseases, arthritis,  
XX CC organ rejection, diabetes and diabetic complications, stroke, ischemia,  
XX CC Alzheimer's disease, myocardial infarction, haemorrhagic shock, diabetic  
XX CC retinopathy, diabetes, venous insufficiency, unstable angina or trauma.  
XX CC They can be used in the veterinary treatment of a non-human subject.  
XX CC Protease inhibitors can be used to lower cell activation resulting from  
XX CC these diseases and deficiencies. The detection of an elevated level of  
XX CC hydrogen peroxide can be used to detect an inflammatory condition. An  
XX CC elevated level of hydrogen peroxide in plasma or whole blood and in the  
XX CC presence of superoxide dismutase (SOD) indicates leukocyte up  
XX CC regulation, e.g. indicative of the onset of an acute cardiovascular  
XX CC disorders, such as disease onset or ischemic complications. An elevated  
XX CC level of hydrogen peroxide in plasma or whole blood and a low level in  
XX CC the presence of SOD is indicative of a chronic or immune compromised  
XX CC condition e.g. hypertension or sepsis. AAY50201-Y50334 represent peptides  
XX CC used in the method of the invention.  
XX SQ Sequence 11 AA;  
Query Match 61.8%; Score 42; DB 20; Length 11;  
Best Local Similarity 72.7%; Pred. No. 1.2;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CGVPAIQPVLS 11  
DB 1 CGVPSIPPNLS 11  
RESULT 14  
ABGL4637  
ID ABGL4637 standard; Protein; 107 AA.  
XX AC ABGL4637;  
XX DT 18-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #14628.  
XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX KW

XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PP 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX PI WPI; 2001-639362/73.  
XX DR N-PSDB; AAS78824.  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity -  
XX PS Claim 20; SEQ ID No 44996; 103pp; English.  
XX CC The invention relates to isolated polynucleotide (I) and  
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX CC and gene mapping, and in recombinant production of (II). The  
XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques  
XX CC to restore normal activity of (II) or to treat disease states involving  
XX CC a quantitating a polypeptide in tissue, as molecular weight markers and as  
XX CC a food supplement. (II) and its binding partners are useful in medical  
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
XX CC disorders involving aberrant protein expression or biological activity.  
XX CC The polypeptide and polynucleotide sequences have applications in  
XX CC diagnostics, forensics, gene mapping, identification of mutations  
XX CC responsible for genetic disorders or other traits to assess biodiversity  
XX CC and to produce other types of data and products dependent on DNA and  
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human  
XX CC diagnostic amino acid sequences of the invention.  
XX CC Note that sequence data for this patent did not appear in the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 107 AA;  
Query Match 61.8%; Score 42; DB 22; Length 107;  
Best Local Similarity 66.7%; Pred. No. 15;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 2 GVPATQPVLSGL 13  
DB 11 GDELPQVPLAGL 22  
RESULT 15  
AAU68565  
ID AAU68565 standard; Protein; 107 AA.  
XX AC AAU68565;  
XX DT 16-JAN-2002 (first entry)  
XX DE Human novel cytokine encoded by cDNA 790CIP2A\_11 #2.  
XX DE Human; cytokine; cell proliferation; cell differentiation;  
XX KW antinflammatory; stem cell growth factor; activin; inhibin; cancer;  
XX KW nervous system disease; neuropathy; Alzheimer's disease;  
XX KW Parkinson's disease; Huntington's disease; spinal cord disorder;  
XX KW

KM	head trauma; stroke; myeloid cell disorder; lymphoid cell disorder;
KM	platelet disorder; thrombocytopaenia; stem cell disorder;
KM	aplastic anaemia; tissue regeneration; wound healing; ulcer;
KM	osteoporosis; osteoarthritis; bone degenerative disorder;
KM	peridontal disease; fibrosis; reparfusion; immune disorder; SCID;
KM	severe combined immunodeficiency; infection; autoimmune disorder;
KM	multiple sclerosis; rheumatoid arthritis; diabetes mellitus; allergy;
KM	asthma; coagulation disorder; haemophilia; sepsis; nephritis;
KM	inflammatory bowel disease; food supplement; immunogen.
XX	

OS Homo sapiens.

XX  
PN WO200175093-A1.

XX  
PD 11-OCT-2001

XX  
PF  
30-MAR-2001:

XX  
PR 31-MAR-2000: 2000US-0540217

PR 23-AUG-2000; 2000US-0649167.  
PR 22-SEP-2000 2000US-0558580

PR 23-OCT-2000; 2000US-0695618.  
PR 30-NOV-2000 2000US 0328711

PR 14-MAR-2001; 2000US-0728711.  
XX

PA (HYSE-) HYSEQ INC.  
YY

PI Tang YT, Asundi V, Zhou P,

PI Yang Y, Zhao QA, Chen R, W  
XX  
22

DR WPI; 2001-626432/7.  
DR N-PSDB; AAS59857.

XX  
PT  
New polypeptides anPT inflammatory, auto-  
PT degenerative disorderXX  
PS Claim 20; Page 308

XX The invention relates

CC cytokines) and the  
CC for identifying a

CC and antagonists). The prob  
CC detection. The prob

CC sources or supplements  
CC to cytokine, cell

stem cell growth factors, suppressing and activating

antibodies raised against the diagnosis and the

central and peripheral  
Alzheimer's patients

lateral sclerosis,

diseases, stroke, thrombocytopenia,

regeneration of bone growth, and in tissue

creating osteoporosis  
periodontal disease

various tissues, various immune deficiencies and disorders including severe combined immunodeficiency (SCID), bacterial or fungal infections, autoimmune disorders (e.g. multiple sclerosis, rheumatoid arthritis, such as mellitus, myasthenia gravis), allergic reactions and conditions, other respiratory problems, coagulation disorders, haemophilias, viral shock, sepsis, arthritis, nephritis and inflammatory bowel disease. The present sequence is useful in altering bodily characteristics. The present sequence represents a novel protein of the invention.

IX  
Sequence 107 AA:

Query Match

Query Match	Score 42;	DB 22;	Length 107;
Best Local Similarity	61.8%	66.7%	66.7%

Best Local Similarity 66.7%; Pred. No. 15;

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:17:50 ; Search time 3.2597 Seconds  
(without alignments)  
383.393 Million cell updates/sec

Title: US-10-036-371-5

Perfect score: 68

Sequence: 1 CGVPAIQPVLSGL 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 73:\*

1: Pir1:\*

2: Pir2:\*

3: Pir3:\*

4: Pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	100.0	245	1 KYBOB	chymotrypsin (EC 3
2	68	100.0	245	1 KYBOA	chymotrypsin (EC 3
3	68	100.0	263	2 A21195	chymotrypsin (EC 3
4	63	92.6	263	2 A31299	chymotrypsin (EC 3
5	62	91.2	16	2 C61414	chymotrypsin (EC 3
6	62	91.2	17	2 B61414	chymotrypsin (EC 3
7	61	89.7	263	1 KYRTB	chymotrypsin (EC 3
8	58	85.3	28	2 A61529	chymotrypsin (EC 3
9	54	79.4	20	2 A61414	chymotrypsin (EC 3
10	53	77.9	25	2 A45109	chymotrypsin (EC 3
11	49	72.1	264	2 I38136	chymotrypsin (EC 3
12	47	69.1	263	2 A23473	chymotrypsin-like
13	46	67.6	126	2 A23473	chymotrypsin-like
14	46	67.6	244	2 S72219	chymotrypsin-like
15	44.5	65.4	384	2 AH0241	chymotrypsin B - A
16	41	60.3	342	2 AD3450	ceramide glucosylt
17	40	58.8	255	2 B87595	proline racemase (
18	40	58.8	475	1 RKQML	transcription regu
19	39.5	58.1	364	2 S44899	ribulose-bisphosph
20	39.5	58.1	396	1 TRYXBA	ZK1236.4 protein -
21	39	57.4	86	2 AC2549	alpha-lytic protei
22	39	57.4	241	2 C81971	hypothetical prote
23	39	57.4	241	2 H81026	hypothetical prote
24	39	57.4	282	2 A84341	conserved hypothet
25	39	57.4	523	1 A44195	hypothetical prote
26	39	57.4	910	2 A01796	Uf21 protein homol
27	38	55.9	111	2 C71227	NADH2 dehydrogenas
28	38	55.9	206	2 H47021	hypothetical prote
29	38	55.9	268	2 S68826	pectic enzyme secr
					pancreatic elastas

30	38	55.9	268	2 S68825	pancreatic elastas
31	38	55.9	269	2 C70710	hypothetical prote
32	38	55.9	400	2 A10261	probable multidrug
33	38	55.9	452	2 C82099	conserved hypothet
34	38	55.9	475	1 RKITL	ribulose-bisphosph
35	38	55.9	559	2 T02825	probable membrane
36	38	55.9	829	2 T19494	hypothetical prote
37	38	55.9	1048	1 XPBEA9	large structural p
38	37.5	55.1	226	2 T35172	hypothetical prote
39	37.5	55.1	237	2 A70672	hypothetical prote
40	37.5	55.1	345	2 T32416	hypothetical prote
41	37.5	55.1	466	1 TWFF	transcription fact
42	37.5	55.1	525	2 T25550	hypothetical prote
43	37.5	55.1	557	2 S21596	extracellular prote
44	37	54.4	126	2 G84399	hypothetical prote
45	37	54.4	378	2 T19012	hypothetical prote

#### ALIGNMENTS

RESULT 1  
KYBOB  
chymotrypsin (EC 3.4.21.1) B precursor - bovine  
N/Alternate names: chymotrypsinogen B  
C/Species: Bos primigenius taurus (cattle)  
C/Date: 08-Oct-1981 #sequence\_revision 08-Oct-1981 #text\_change 18-Jul-1997  
C/Accession: A00953  
R/Smillie, L.B.; Furka, A.; Nagabhushan, N.; Stevenson, K.J.; Parkes, C.O.  
A/Title: Structure of chymotrypsinogen B compared with chymotrypsinogen A and trypsinogen  
A/Reference number: A00953; MUID:68238908; PMID:5649671  
A/Accession: A00953  
A/Molecule type: protein  
A/Residues: 1-245 <SMI>  
C/Comment: Chymotrypsinogen B is synthesized, along with chymotrypsinogen A, in the acin.  
C/Comment: The first activation cleavage, leading to pi-chymotrypsin B, occurs in the sar  
C/Superfamily: trypsin; trypsin homology  
C/Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen  
F.1-15/Domain: propeptide #status experimental <PRO>  
F.16-245/Product: chymotrypsin B #status experimental <TRY>  
F.16-238/Domain: trypsin homology <TRY>  
F.1-122,42-58,136-201,168-182,191-220/Disulfide bonds: #status experimental  
F.57,102,195/Active site: His, Asp, Ser #status experimental

Query Match 100.0% Score 68; DB 1; Length 245;  
Best Local Similarity 100.0%; Pred. NO. 0.00057;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGVPAIQPVLSGL 13  
Db 1 CGVPAIQPVLSGL 13

#### RESULT 2

KYBOA  
chymotrypsin (EC 3.4.21.1) A precursor - bovine  
N/Alternate names: chymotrypsinogen A  
C/Species: Bos primigenius taurus (cattle)  
C/Date: 07-May-1981 #sequence\_revision 07-May-1981 #text\_change 07-May-1999  
C/Accession: A90235; A93158; S29650; A00952  
R/Brown, J.R.; Hartley, B.S.  
Biochem. J. 101, 214-228, 1966  
A/Title: Location of disulphide bridges by diagonal paper electrophoresis. The disulphid  
A/Reference number: A90235; MUID:67181721; PMID:5971783  
A/Accession: A90235  
A/Molecule type: protein  
A/Residues: 1-101, N<sup>103</sup>-245 <BRO>  
R/Blow, D.M.; Birktoft, J.J.; Hartley, B.S.  
Nature 221, 337-340, 1969  
A/Title: Role of a buried acid group in the mechanism of action of chymotrypsin.  
A/Reference number: A93158; MUID:69106266; PMID:5764436  
A/Contents: annotation; revision to residue 102

Wed Feb 12 11:59:40 2003

R;Meloun, B.; Kluh, I.; Kostka, V.; Moravek, L.; Prusik, Z.; Vanacek, J.; Keil, B.; Som  
Biochim. Biophys. Acta 130, 543-546, 1966  
A;Title: Covalent structure of bovine chymotrypsinogen A.  
A;Reference number: A90572; MUID:67183948; PMID:5972866  
A;Accession: A93158  
A;Molecule type: protein  
A;Residues: 1-101, 'N', 103-245 <MEL>  
A;Note: disulfide bonds were determined  
R;Cutruzzola, P.; Ascenzi, P.; Barra, D.; Bolognesi, M.; Menegatti, E.; Sarti, P.; Schne  
Biochim. Biophys. Acta 1161, 201-208, 1993  
A;Title: Selective oxidation of Met-192 in bovine alpha-chymotrypsin. Effect on catalytic  
A;Reference number: S29650; MUID:93160238; PMID:8431470  
A;Accession: S29650  
A;Molecule type: protein  
A;Residues: 1-12,16-27;149-160;181-200 <CUT>  
R;Smillie, L.B.; Hartley, B.S.  
Biochem. J. 101, 232-241, 1966  
A;Title: Histidine sequences in the active centres of some 'serine' proteinases.  
A;Reference number: A90336; MUID:67181723; PMID:5971785  
A;Contents: annotation; active site  
R;Birktoft, J.J.; Blow, D.M.; Henderson, R.; Steitz, T.A.  
Philos. Trans. R. Soc. Lond. B257, 67-76, 1970  
A;Title: The structure of alpha-chymotrypsin.  
A;Reference number: A93754  
A;Contents: annotation; X-ray crystallography  
A;Comment: Chymotrypsinogens are synthesized in the acinar cells of pancreas.  
C;Comment: Tryptic cleavage after Arg-15 results in a fully active enzyme (pi-chymotryps  
delta-chymotrypsin; further chymotryptic cleavage liberates the dipeptide Thr-147 and A  
d Aen-148 directly from chymotrypsinogen, which leads to the degraded form neochymotryp  
C;Superfamily: trypsin; trypsin homology  
C;Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen  
F;1-13,16-146,149-245/Product: alpha-chymotrypsin #status experimental <ZYM>  
F;1-13,16-146,149-245/Product: alpha-chymotrypsin #status experimental <MPT>  
F;16-238/Domain: trypsin homology <TRY>  
F;1-122,42-58,136-201,168-182,191-220/Disulfide bonds: #status experimental  
F;57,102,195/Active site: His, Asp, Ser #status experimental

Query Match 100.0%; Score 68; DB 1; Length 245;  
Best Local Similarity 100.0%; Pred. No. 0.00057; Mismatches 0; Indels 0; Gaps 0;  
Matches 13; Conservative 0;

Qy 1 CGVPAIQPVL SGL 13  
Db 1 CGVPAIQPVL SGL 13  
|||||

RESULT 3  
A21195  
chymotrypsin (EC 3.4.21.1) 2 precursor - dog  
C;Species: Canis lupus familiaris (dog)  
C;Date: 27-Jul-1990 #sequence\_revision 27-Jul-1990 #text\_change 22-Jun-1999  
C;Accession: A21195  
R;Pinaky, S.D.; LaForge, K.S.; Luc, V.; Scheele, G.  
Proc. Natl. Acad. Sci. U.S.A. 80, 7486-7490, 1983  
A;Title: Identification of cDNA clones encoding secretory isoenzyme forms: sequence dete  
A;Reference number: A21195; MUID:84170253; PMID:6584866  
A;Accession: A21195  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-263 <PIN>  
A;Cross-references: GB:K01173; NID:g163945; PIDN:AAA30841.1; PID:g163946  
C;Superfamily: trypsin; trypsin homology  
C;Keywords: hydrolase; protein digestion; serine proteinase  
F;34-236/Domain: trypsin homology <TRY>  
F;75,120,213/Active site: His, Asp, Ser #status predicted

Query Match 100.0%; Score 68; DB 2; Length 263;  
Best Local Similarity 100.0%; Pred. No. 0.00061; Mismatches 0; Indels 0; Gaps 0;  
Matches 13; Conservative 0;

Qy 1 CGVPAIQPVL SGL 13  
Db 19 CGVPAIQPVL SGL 31  
|||||

RESULT 4  
A31299  
chymotrypsin (EC 3.4.21.1) precursor - human  
C;Species: Homo sapiens (man)  
C;Date: 08-Jun-1989 #sequence\_revision 08-Jun-1989 #text\_change 22-Jun-1999  
C;Accession: A31299  
R;Tomita, N.; Izumoto, Y.; Horii, A.; Doi, S.; Yokouchi, H.; Ogawa, M.; Mori, T.; Matsub  
Biochem. Biophys. Res. Commun. 158, 569-575, 1989  
A;Title: Molecular cloning and nucleotide sequence of human pancreatic prechymotrypsinoge  
A;Reference number: A31299; MUID:89134264; PMID:2917002  
A;Accession: A31299  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-263 <TOM>  
A;Cross-references: GB:M24400; NID:g181189; PIDN:AAA52128.1; PID:g181190  
C;Genetics:  
A;Gene: GDB:CTRB1; CTRB  
A;Cross-references: GDB:119820; OMIM:118890  
A;Map position: 16q23.1-16q23.1  
C;Superfamily: trypsin; trypsin homology  
C;Keywords: hydrolase; protein digestion; serine proteinase  
F;34-236/Domain: trypsin homology <TRY>  
F;75,120,213/Active site: His, Asp, Ser #status predicted

Query Match 92.6%; Score 63; DB 2; Length 263;  
Best Local Similarity 92.3%; Pred. No. 0.0043; Mismatches 0; Indels 1; Gaps 0;  
Matches 12; Conservative 0;

Qy 1 CGVPAIQPVL SGL 13  
Db 19 CGVPAIQPVL SGL 31  
|||||

RESULT 5  
C61414  
chymotrypsin (EC 3.4.21.1) - slider turtle (fragment)  
C;Species: Pseudemys scripta (slider)  
C;Date: 09-Sep-1994 #sequence\_revision 09-Sep-1994 #text\_change 07-May-1999  
C;Accession: C61414  
R;Bhargava, A.K.; Barnard, E.A.  
J. Mol. Evol. 2, 187-198, 1973  
A;Title: Evolution in the pancreatic chymotrypsinogen series: N-terminal sequence determ  
A;Reference number: A61414; MUID:76146602; PMID:4807189  
A;Accession: C61414  
A;Status: preliminary  
A;Molecule type: protein  
A;Residues: 1-16 <BHA>  
C;Keywords: hydrolase; protein digestion; serine proteinase

Query Match 91.2%; Score 62; DB 2; Length 16;  
Best Local Similarity 92.3%; Pred. No. 0.00042; Mismatches 0; Indels 1; Gaps 0;  
Matches 12; Conservative 0;

Qy 1 CGVPAIQPVL SGL 13  
Db 1 CGVPAIQPVL SGL 13  
|||||

RESULT 6  
B61414  
chymotrypsin (EC 3.4.21.1) - painted turtle (fragment)  
C;Species: Chrysemys picta (painted turtle)  
C;Date: 19-Mar-1997 #sequence\_revision 19-Dec-1997 #text\_change 07-May-1999  
C;Accession: B61414  
R;Bhargava, A.K.; Barnard, E.A.  
J. Mol. Evol. 2, 187-198, 1973  
A;Title: Evolution in the pancreatic chymotrypsinogen series: N-terminal sequence determ  
A;Reference number: A61414; MUID:76146602; PMID:4807189  
A;Accession: B61414  
A;Status: preliminary  
A;Molecule type: protein

```

A;Residues: 1-17 <BHA>
C;Keywords: hydrolase; serine proteinase

Query Match      91.2%; Score 62; DB 2; Length 17;
Best Local Similarity 92.3%; Pred. No. 0.00044;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLSGL 13
    ||||| |||||
Db 1 CGVPAIPVLSGL 13

RESULT 7
KYRTB
chymotrypsin (EC 3.4.21.1) B precursor - rat
N;Alternate names: chymotrypsinogen B
C;Species: Rattus norvegicus (Norway rat)
C;Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 18-Jun-1999
C;Accession: A22658
R;Bell, G.I.; Quinto, C.; Quiroga, M.; Valenzuela, P.; Craik, C.S.; Rutter, W.J.
J. Biol. Chem. 259, 14265-14270, 1984
A;Title: Isolation and sequence of a rat chymotrypsin B gene.
A;Reference number: A22658; MUID:85054881; PMID:6209274
A;Accession: A22658
A;Molecule type: DNA
A;Residues: 1-263 <BEL>
A;Cross-references: GB:K02298; NID:g203653; PIDN:AAA98732.1; PID:g203654
C;Genetics:
A;Introns: 18/1; 52/3; 79/2; 105/3; 166/1; 210/3
C;Superfamily: trypsin; trypsin homology
C;Keywords: hydrolase; pancreas; protein digestion; serine proteinase
F;19-33/Domain: signal sequence #status predicted <SIG>
F;34-263/Domain: propeptide #status predicted <PRO>
F;34-263/Product: chymotrypsin B #status predicted <MAT>
F;34-266/Domain: trypsin homology <TRY>
F;75,120,213/Active site: His, Asp, Ser #status predicted

Query Match      89.7%; Score 61; DB 1; Length 263;
Best Local Similarity 84.6%; Pred. No. 0.0093;
Matches 11; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLSGL 13
    ||||| |||||
Db 19 CGVPTIQPVLTGL 31

RESULT 8
A61529
chymotrypsin (EC 3.4.21.1) - Atlantic cod (fragments)
C;Species: Gadus morhua (Atlantic cod)
C;Date: 15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change 11-May-2000
C;Accession: A61529
R;Asgeirsson, B.; Bjarnason, J.B.
Comp. Biochem. Physiol. B 99, 327-335, 1991
A;Title: Structural and kinetic properties of chymotrypsin from Atlantic cod (Gadus morhua)
A;Reference number: A61529; MUID:92111252; PMID:1764912
C;Accession: A61529
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-28 <ASG>
C;Superfamily: trypsin; trypsin homology
C;Keywords: hydrolase; protein digestion; serine proteinase

Query Match      85.3%; Score 58; DB 2; Length 28;
Best Local Similarity 76.9%; Pred. No. 0.0034;
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLSGL 13
    ||||| |||||
Db 1 CGSPAIPVLSGL 13

RESULT 9
chymotrypsin (EC 3.4.21.1) - snapping turtle (fragment)
C;Species: Chelydra serpentina (snapping turtle)
C;Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 07-May-1999
C;Accession: A61414
R;Bhargava, A.K.; Barnard, E.A.
J. Mol. Evol. 2, 187-198, 1973
A;Title: Evolution in the pancreatic chymotrypsinogen series: N-terminal sequence determination
A;Reference number: A61414; MUID:76146602; PMID:4807189
A;Accession: A61414
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-20 <BHA>
C;Keywords: hydrolase; protein digestion; serine proteinase

Query Match      79.4%; Score 54; DB 2; Length 20;
Best Local Similarity 84.6%; Pred. No. 0.012;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLSGL 13
    ||||| |||||
Db 1 CVVPIIQPVLSGL 13

RESULT 10
A45109
chymotrypsin (EC 3.4.21.1) homolog p31 - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 31-Mar-1997
C;Accession: A45109
R;Gorr, S.U.; Hamilton, J.W.; Cohn, D.V.
J. Biol. Chem. 267, 21595-21600, 1992
A;Title: Regulated, but not constitutive, secretory proteins bind porcine chymotrypsinogen
A;Reference number: A45109; MUID:93016107; PMID:1400470
A;Accession: A45109
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-25 <GOR>
A;Experimental source: pancreas
A;Note: sequence extracted from NCBI backbone (NCBIP:116723)
C;Superfamily: trypsin; trypsin homology
C;Keywords: hydrolase; serine proteinase

Query Match      77.9%; Score 53; DB 2; Length 25;
Best Local Similarity 91.7%; Pred. No. 0.021;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GVPPIQPVLSGL 13
    ||||| |||||
Db 2 GVPPIPPVLSGL 13

RESULT 11
I38136
chymotrypsin-like proteinase (EC 3.4.21.-) CTRL-1 - human
C;Species: Homo sapiens (man)
C;Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 22-Jun-1999
C;Accession: I38136
R;Larsen, F.; Solheim, J.; Kristensen, T.; Kolsto, A.B.; Prydz, H.
Hum. Mol. Genet. 2, 1589-1595, 1993
A;Title: A tight cluster of five unrelated human genes on chromosome 16q22.1.
A;Reference number: I38135; MUID:9409344; PMID:8268911
A;Accession: I38136
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-264 <RES>
A;Cross-references: EMBL:X71874; NID:g406226; PIDN:CAA50710.1; PID:g406228
C;Genetics:
A;Gene: GDB:CTRL
A;Cross-references: GDB:204061
A;Map position: 16q22.1-16q22.1
A;Introns: 18/1; 52/3; 79/2; 106/3; 167/1; 211/3
C;Superfamily: trypsin; trypsin homology

```

Wed Feb 12 11:59:40 2003

us-10-036-371-5.rpx

C;Keywords: hydrolase; serine proteinase  
F;34-257/Domain: trypsin homology <TRY>  
F;75,121,214/Active site: His, Asp, Ser #status predicted

Query Match 72.1%; Score 49; DB 2; Length 264;  
Best Local Similarity 72.7%; Pred. No. 1;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLS 11  
|||:|:|:|:|  
Db 19 CGIPAIPKPAUS 29

## RESULT 12

S47537 chymotrypsin (EC 3.4.21.1) precursor - Atlantic cod

C;Species: Gadus morhua (Atlantic cod)  
C;Date: 26-Dec-1994 #sequence\_revision 03-Aug-1995 #text\_change 22-Jun-1999  
C;Accession: S47537; S43163  
R;Gudmundsdottir, A.; Oskarsen, S.; Eakin, A.E.; Craik, C.S.; Bjarnason, J.B.  
Biochim. Biophys. Acta 1219, 211-214, 1994  
A;Title: Atlantic cod cDNA encoding a psychrophilic chymotrypsinogen.  
A;Reference number: S47537; MUID:94368860; PMID:8086467  
A;Accession: S47537  
A;Molecule type: mRNA  
A;Residues: 1-263 <GUD>  
A;Cross-references: EMBL:X78490; NID:G468750; PIDN:CAA55242.1; PID:G468751  
C;Superfamily: trypsin; trypsin homology  
C;Keywords: hydrolase; protein digestion; serine proteinase  
F;1-18/Domain: signal sequence #status predicted <SIG>  
F;19-263/Product: chymotrypsin #status predicted <WAT>  
F;34-256/Domain: trypsin homology <TRY>  
F;75,120,213/Active site: His, Asp, Ser #status predicted

Query Match 69.1%; Score 47; DB 2; Length 263;  
Best Local Similarity 66.7%; Pred. No. 2.2;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLSG 12  
|||:|:|:|:|  
Db 19 CGIPAIPKPAUS 30

## RESULT 13

A23473 chymotrypsin-like proteinase (EC 3.4.21.-) - pig (tentative sequence) (fragments)  
N;Alternate names: pancreatic elastase II [misidentification]  
C;Species: Sus scrofa domestica (domestic pig)  
C;Date: 05-Jun-1987 #sequence\_revision 05-Jun-1987 #text\_change 31-Mar-2000

C;Accession: A23473  
R;Vered, M.; Gertler, A.; Burstein, Y.  
Int. J. Pept. Protein Res. 27, 183-190, 1986  
A;Reference number: A23473; MUID:86194934; PMID:3634756

A;Accession: A23473  
A;Molecule type: protein  
A;Residues: 1-126 <VER>  
C;Superfamily: trypsin; trypsin homology  
C;Keywords: hydrolase; serine proteinase

Query Match 67.6%; Score 46; DB 2; Length 126;  
Best Local Similarity 80.0%; Pred. No. 1.6;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGVPAIQPVL 10  
|||:|:|:|  
Db 1 CGVPAIKPAL 10

## RESULT 14

S72219 chymotrypsin B - Atlantic cod (fragments)

C;Species: Gadus morhua (Atlantic cod)  
C;Date: 19-Mar-1998 #sequence\_revision 17-Apr-1998 #text\_change 07-Aug-1998

C;Accession: S72219  
R;Lech-Larsen, R.; Asgeirsson, B.; Thorolfsson, M.; Norregaard-Madsen, M.; Hojrup, P.  
Biochim. Biophys. Acta 1297, 49-56, 1996  
A;Title: Structure of chymotrypsin variant B from Atlantic cod, Gadus morhua.  
A;Reference number: S72219; MUID:96439045; PMID:8841380

A;Accession: S72219  
A;Status: preliminary  
A;Molecule type: protein  
A;Residues: 1-14;15-244 <LET>  
C;Superfamily: trypsin; trypsin homology  
F;15-237/Domain: trypsin homology <TRY>

Query Match 67.6%; Score 46; DB 2; Length 244;  
Best Local Similarity 66.7%; Pred. No. 3;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLSG 12  
|||:|:|:|:|  
Db 1 CGSPAIPQVPTG 12

## RESULT 15

AH0241 ceramide glucosyltransferase (EC 2.4.1.80) [imported] - Yersinia pestis (strain CO92)  
C;Species: Yersinia pestis  
C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 27-Nov-2001  
C;Accession: AH0241

R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, T.  
Nature 413, 523-527, 2001

A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A;Reference number: AB0001; MUID:21470413; PMID:11586360

A;Accession: AH0241  
A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-384 <KUR>

A;Cross-references: GB:AL590842; PIDN:CAC90796.1; PID:G15979996; GSPDB:GN00175

C;Genetics:

A;Gene: YPO1983

C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 65.4%; Score 44.5; DB 2; Length 384;  
Best Local Similarity 58.8%; Pred. No. 8.2;  
Matches 10; Conservative 1; Mismatches 1; Indels 5; Gaps 1;

QY 1 CGVP-----AIQPVLSG 12  
|||:|:|:|:|  
Db 31 CGVPAQATATILQPVLSG 47

Search completed: February 12, 2003, 10:28:51  
Job time : 5.2597 sec



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:04:45 ; Search time 1.74627 Seconds  
(without alignments)  
308.768 Million cell updates/sec

Title: US-10-036-371-5  
Perfect score: 68  
Sequence: 1 CGVPAIQPVLSGL 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match	Length	ID	Description	
1	68	100.0	245	1	CTRA_BOVIN	P00766 bos taurus
2	68	100.0	245	1	CTRB_BOVIN	P00767 bos taurus
3	68	100.0	263	1	CTR2_CANFA	P04813 canis familiaris
4	63	92.6	263	1	CTRB_HUMAN	P17538 homo sapiens
5	61	89.7	263	1	CTRB_RAT	P07338 rattus norvegicus
6	49	72.1	264	1	CTRL_HUMAN	P40313 homo sapiens
7	47	69.1	263	1	CTRA_GADMO	P47796 gadus morhua
8	46	67.6	245	1	CTRB_GADMO	P06446 gadus morhua
9	40	58.8	475	1	RBL_CHLRE	P00877 chlamydomonas
10	40	58.8	591	1	HE_HEMPU	P19153 hemicentrotus
11	39.5	58.1	364	1	Y084_CAEEL	P14620 caenorhabditis
12	39.5	58.1	397	1	PRLA_LYSEN	P00778 lysochacter
13	39	57.4	282	1	PNK_HAUN1	P09hx7 halobacter
14	39	57.4	523	1	UL21_PRVN3	P00703 pseudorabies
15	39	57.4	907	1	NIUG_SALTY	P13900 salmonella
16	38	55.9	206	1	GSFJ_ERWCH	P24689 erwinia chrysanthemi
17	38	55.9	268	1	CLCR_HUMAN	P09895 homo sapiens
18	38	55.9	294	1	NKZE_CHICK	P00788 gallus gallus
19	38	55.9	452	1	YM53_VIBCH	P09kp9 vibrio cholerae
20	38	55.9	475	1	RBL_ASTLO	P26490 astrasia lon
21	38	55.9	626	1	DALY_DROME	Q24114 drosophila
22	38	55.9	733	1	YA34_HUMAN	Q9upw6 homo sapiens
23	38	55.9	1048	1	P100_HCMVA	P00318 human cytochrome
24	37.5	55.1	466	1	KRUP_DROME	P07247 drosophila
25	37	54.4	269	1	EL2_BOVIN	P02941 bos taurus
26	37	54.4	436	1	RBL_EUGVI	P45481 mus musculus
27	37	54.4	2441	1	CBP_MOUSE	P33892 saccharomyces
28	37	54.4	2672	1	GCNI_YEAST	O02720 canis familiaris
29	36	52.9	167	1	OB_CANFA	Q9czm2 mus musculus
30	36	52.9	214	1	RL15_MOUSE	P12360 lycopersicon
31	36	52.9	246	1	CB11_LYCES	P39197 paracoccus
32	36	52.9	318	1	PTA_PARDE	Q39617 chlamydomonas
33	36	52.9	397	1	POR_CHLRE	

34 36 52.9 429 1 YDIS\_ECOLI  
35 36 52.9 444 1 EXD4\_MOUSE  
36 36 52.9 451 1 ECFE\_YERPE  
37 36 52.9 746 1 CLC5\_HUMAN  
38 36 52.9 746 1 CLC5\_MOUSE  
39 36 52.9 746 1 CLC5\_RAT  
40 36 52.9 747 1 CLC4\_MOUSE  
41 36 52.9 760 1 CLC4\_RAT  
42 36 52.9 760 1 CLC3\_CAVPO  
43 36 52.9 760 1 CLC3\_MOUSE  
44 36 52.9 760 1 CLC3\_RAT  
45 36 52.9 760 1 CLC4\_HUMAN

## ALIGNMENTS

RESULT 1

CTRA\_BOVIN  
ID CTRA\_BOVIN STANDARD; PRT; 245 AA.  
AC P00766;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DE 15-JUN-2002 (Rel. 41, Last annotation update)  
Chymotrypsinogen A (EC 3.4.21.1).  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE, DISULFIDE BONDS, AND ACTIVE SITE.  
RX MEDLINE=67181721; PubMed=5971785;  
RA Brown J.R., Hartley B.S.;  
RT "Location of disulphide bridges by diagonal paper electrophoresis. The disulphide bridges of bovine chymotrypsinogen A.";  
RL Biochem. J. 101:214-228(1966).  
RN [2]  
RP REVISION TO 102.  
RX MEDLINE=69106266; PubMed=5764436;  
RA Blow D.M., Birktoft J.J., Hartley B.S.;  
RT "Role of a buried acid group in the mechanism of action of chymotrypsin.";  
RL Nature 221:337-340(1969).  
RN [3]  
RP PRELIMINARY SEQUENCE.  
RA Hartley B.S.;  
RT "Amino-acid sequence of bovine chymotrypsinogen-A.";  
RL Nature 201:1284-1287(1964).  
RN [4]  
RP SEQUENCE, AND DISULFIDE BONDS.  
RX MEDLINE=67183948; PubMed=5972866;  
RA Meloun B., Klueh I., Kostka V., Moravsek L., Prusik Z., Vanacek J., Keil B., Sorm F.;  
RT "Covalent structure of bovine chymotrypsinogen A.";  
RL Biochim. Biophys. Acta 130:543-546(1966).  
RN [5]  
RP ACTIVE SITE.  
RX MEDLINE=67181723; PubMed=5971785;  
RA Smillie L.B., Hartley B.S.;  
RT "Histidine sequences in the active centres of some 'serine' proteinases.";  
RL Biochem. J. 101:232-241(1966).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY.  
RX MEDLINE=72035052; PubMed=4399050;  
RA Birktoft J.J., Blow D.M., Henderson R., Steitz T.A.;  
RT "I. Serine proteinases. The structure of alpha-chymotrypsin.";  
RL Philos. Trans. R. Soc. Lond. B, Biol. Sci. 257:67-76(1970).  
RN [7]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF CHYMOTRYPSINOGEN.  
RX MEDLINE=70177557; PubMed=5442169;  
RA Freer S.T., Kraut J., Robertus J.D., Wright H.T., Xuong N.H.;

CHAIN	149	245	CHYMOTRYPSIN A, C CHAIN.
ACT_SITE	57	57	CHARGE RELAY SYSTEM.
ACT_SITE	102	102	CHARGE RELAY SYSTEM.
ACT_SITE	195	195	CHARGE RELAY SYSTEM.
DISULFID	1	122	
DISULFID	42	58	
DISULFID	136	201	
DISULFID	168	182	
DISULFID	191	220	
HELIX	12	15	
STRAND	20	21	
TURN	24	25	
TURN	28	29	
STRAND	30	34	
STRAND	40	46	
STRAND	51	54	
TURN	57	58	
TURN	62	63	
STRAND	65	68	
TURN	69	69	
TURN	72	72	
TURN	73	74	
STRAND	81	90	
STRAND	92	93	
STRAND	95	95	
TURN	96	99	
STRAND	100	100	
TURN	101	101	
STRAND	104	108	
STRAND	122	122	
TURN	126	127	
TURN	132	133	
STRAND	135	140	
TURN	146	148	
STRAND	154	154	
STRAND	156	163	
HELIX	165	168	
TURN	169	172	
HELIX	173	175	
STRAND	180	184	
TURN	192	193	
TURN	195	196	
STRAND	198	203	
TURN	204	205	
STRAND	206	216	
TURN	218	219	
TURN	222	223	
STRAND	225	230	
HELIX	231	233	
TURN	234	234	
HELIX	235	243	
SEQUENCE	245 AA;	25666 MW;	91A9F28E2F3E3142 CRC64;
Query Match	100.0%;	Score 68;	DB 1;
Best Local Similarity	100.0%;	Pred. NO.	0.00019;
Matches 13;	Conservative 0;	Mismatches 0;	Indels 0;
Qy	1	CGVPAIQPVLSGL 13	Gaps 0;
Db	1	CGVPAIQPVLSGL 13	
RESULT 2			
ID	CTRB_BOVIN	STANDARD;	PRT; 245 AA.
AC	P00767;		
DT	21-JUL-1986 (Rel. 01, Created)		
DT	21-JUL-1986 (Rel. 01, Last sequence update)		
DE	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Chymotrypsinogen B (EC 3.4.21.1).		
OS	Bos taurus (Bovine).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		

```

CC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE, DISULFIDE BONDS, AND ACTIVE SITE.
RX MEDLINE=68238908; PubMed=5649671;
RA "Smillie L.B., Furka A., Nagabhushan N., Stevenson K.J., Parkes C.O.;
RT "Structure of chymotrypsinogen B compared with chymotrypsinogen A and
RL tryptelogenen";
RL Nature 218:343-346(1968).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
CC Phe-|-Xaa, Leu-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- DATABASE: NAME=Morthington enzyme manual;
CC WWW="http://www.worthington-biochem.com/manual/c/CHY.html".
DR PIR; AG0953; KYBOB.
DR HSSP; P00766; IACB.
DR MEROPS; S01152; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SMO0020; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen; Signal.
FT CHAIN 1 13 CHYMOTRYPSIN B, A CHAIN.
FT CHAIN 16 146 CHYMOTRYPSIN B, B CHAIN.
FT CHAIN 149 245 CHYMOTRYPSIN B, C CHAIN.
FT ACT_SITE 57 57 CHARGE RELAY SYSTEM.
FT ACT_SITE 102 102 CHARGE RELAY SYSTEM.
FT ACT_SITE 195 195 CHARGE RELAY SYSTEM.
FT DISULFID 1 122
FT DISULFID 42 58
FT DISULFID 136 201
FT DISULFID 168 182
FT DISULFID 191 220
SQ SEQUENCE 245 AA; 25755 MW; 678016446FF5FEB5 CRC64;

Query Match 100.0%; Score 68; DB 1; Length 245;
Best Local Similarity 100.0%; Pred. No. 0.00019;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLSGL 13
Db 1 CGVPAIQPVLSGL 13

RESULT 3
CTR2 CANFA
ID CTR2 CANFA STANDARD; PRT; 263 AA.
AC P04813;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 13-JUN-2002 (Rel. 41, Last annotation update)
DE Chymotrypsinogen 2 precursor (EC 3.4.21.1).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84170253; PubMed=6584866;
RA Pinsky S.D., LaForge K.S., Luc V., Scheele G.;
RT "Identification of cDNA clones encoding secretory isoenzyme forms:
RT sequence determination of canine pancreatic precchymotrypsinogen 2
RT mRNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:7486-7490(1983).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
CC Phe-|-Xaa, Leu-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; K01173; AAA30841.1; -.
DR PIR; A21195; A21195.
DR HSSP; P00766; IACB.
DR MEROPS; S01152; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SMO0020; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen; Signal.
FT CHAIN 1 18 CHYMOTRYPSINOGEN 2.
FT CHAIN 19 263 CHYMOTRYPSIN 2, A CHAIN.
FT CHAIN 19 31 CHYMOTRYPSIN 2, B CHAIN.
FT CHAIN 167 263 CHYMOTRYPSIN 2, C CHAIN.
FT ACT_SITE 75 75 CHARGE RELAY SYSTEM.
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM.
FT ACT_SITE 213 213 CHARGE RELAY SYSTEM.
FT DISULFID 19 140 BY SIMILARITY.
FT DISULFID 60 76 BY SIMILARITY.
FT DISULFID 154 219 BY SIMILARITY.
FT DISULFID 186 200 BY SIMILARITY.
FT DISULFID 209 238 BY SIMILARITY.
SQ SEQUENCE 263 AA; 27787 MW; 2A2F449D813B3961 CRC64;

Query Match 100.0%; Score 68; DB 1; Length 263;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLSGL 13
Db 19 CGVPAIQPVLSGL 31

RESULT 4
CTR8 HUMAN
ID CTR8 HUMAN STANDARD; PRT; 263 AA.
AC P17538;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymotrypsinogen B precursor (EC 3.4.21.1).
CN CTR81 OR CTR8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=89134264; PubMed=2917002;
RA Tomita N., Izumoto Y., Horii A., Doi S., Yokouchi H., Ogawa M.,
RA Mori T., Matsubara K.;
RT "Molecular cloning and nucleotide sequence of human pancreatic
RT precchymotrypsinogen cDNA.";
RL Biochem. Biophys. Res. Commun. 158:569-575(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

```

```

CC CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
CC CC Phe-|-Xaa, Leu-|-Xaa.
CC CC -1- SUBCELLULAR LOCATION: Extracellular.
CC CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC DR EMBL: M24400; AA52128.1; -.
CC DR EMBL: BC005385; AA05385.1; -.
CC DR PIR: A31299; A31299.
CC DR HSP: P00766; ICHG.
CC DR MEROPS: S01.152; -.
CC DR MIM: 118890; -.
CC DR InterPro: IPR001314; Chymotrypsin.
CC DR Pfam: PF00089; trypsin_1.
CC DR PRINTS: PR00722; CHYMOTRYPSIN.
CC DR SMART: SM00020; Tryp_Spc; 1.
CC DR PROSITE: PS00240; TRYPSIN_DOM; 1.
CC DR PROSITE: PS00134; TRYPSIN_HIS; 1.
CC DR PROSITE: PS00135; TRYPSIN_SER; 1.
CC DR Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal.
CC KW SIGNAL 1 18
CC FT CHAIN 19 263 CHYMOTRYPSINOGEN B.
CC FT CHAIN 19 31 CHYMOTRYPSIN B, A CHAIN.
CC FT CHAIN 34 164 CHYMOTRYPSIN B, B CHAIN.
CC FT CHAIN 167 263 CHYMOTRYPSIN B, C CHAIN.
CC FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT DISULFID 60 76 BY SIMILARITY.
CC FT DISULFID 154 219 BY SIMILARITY.
CC FT DISULFID 186 200 BY SIMILARITY.
CC FT DISULFID 209 238 BY SIMILARITY.
CC SQ SEQUENCE 263 AA; 27870 MW; 4C1C055A490B701 CRC64;

Query Match 92.6%; Score 63; DB 1; Length 263;
Best Local Similarity 92.3%; Pred. No. 0.0014;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLSGL 13
DB 19 CGVPAIHPVLSGL 31

RESULT 5
CTRB_RAT
ID_CTRB_RAT STANDARD; PRT; 263 AA.
AC P07338;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymotrypsinogen B precursor (EC 3.4.21.1).
GN CTRB1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85054881; PubMed=6209274;
RA Bell G.I., Quinco C., Quiroga M., Valenzuela P., Craik C.S.,
RA Rutter W.J.;
RA "Isolation and sequence of a rat chymotrypsin B gene.";
RL J. Biol. Chem. 259:14265-14270(1984).

```

```

CC CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
CC CC Phe-|-Xaa, Leu-|-Xaa.
CC CC -1- SUBCELLULAR LOCATION: Extracellular.
CC CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC DR EMBL: K02298; AAA98732.1; -.
CC DR PIR: A26588; KYRB.
CC DR HSP: P00766; ICHG.
CC DR MEROPS: S01.152; -.
CC DR InterPro: IPR001314; Chymotrypsin.
CC DR Pfam: PF00089; trypsin_1.
CC DR PRINTS: PR00722; CHYMOTRYPSIN.
CC DR SMART: SM00020; Tryp_Spc; 1.
CC DR PROSITE: PS00240; TRYPSIN_DOM; 1.
CC DR PROSITE: PS00134; TRYPSIN_HIS; 1.
CC DR PROSITE: PS00135; TRYPSIN_SER; 1.
CC DR Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal.
CC KW SIGNAL 1 18
CC FT CHAIN 19 263 CHYMOTRYPSINOGEN B.
CC FT CHAIN 19 31 CHYMOTRYPSIN B, A CHAIN.
CC FT CHAIN 34 164 CHYMOTRYPSIN B, B CHAIN.
CC FT CHAIN 167 263 CHYMOTRYPSIN B, C CHAIN.
CC FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT DISULFID 60 76 BY SIMILARITY.
CC FT DISULFID 154 219 BY SIMILARITY.
CC FT DISULFID 186 200 BY SIMILARITY.
CC FT DISULFID 209 238 BY SIMILARITY.
CC SQ SEQUENCE 263 AA; 27849 MW; ACAFDACF8C4DA6D CRC64;

Query Match 89.7%; Score 61; DB 1; Length 263;
Best Local Similarity 84.6%; Pred. No. 0.0031;
Matches 11; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLSGL 13
DB 19 CGVPTIQPVLTGL 31

RESULT 6
CTRL_HUMAN
ID_CTRL_HUMAN STANDARD; PRT; 264 AA.
AC P40313;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymotrypsin-like protease CTRL-1 precursor (EC 3.4.21.-).
GN CTRL OR CTRL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94093544; PubMed=8268911;
RA Larsen F., Solheim J., Kristensen T., Kolsto A.B., Prydz H.;
RA "A tight cluster of five unrelated human genes on chromosome
RA 16q22.1."
RL Hum. Mol. Genet. 2:1589-1595(1993).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).

CC -----  
 CC EMBL; X71874; CAA50710.1; -  
 CC EMBL; X71877; CAA50711.1; -  
 CC HSSP; P00763; IDPO.  
 CC MEROPS; S01.256; -  
 CC Genew; HGNC:2524; CTRL.  
 CC MIM; 118888; -  
 CC InterPro; IPR001314; Chymotrypsin.  
 CC InterPro; IPR001254; Ser. protease\_Try.  
 CC Pfam; PF00089; trypsin; 1.  
 CC PRINTS; PR00722; CHYMOTRYPSIN.  
 CC SMART; SM00200; Tryp\_Spc; 1.  
 CC PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
 CC Hydrolase; Serine protease; Zymogen; Signal.  
 CC SIGNAL 1 18  
 CC CHAIN 19 33  
 CC ACT\_SITE 34 264  
 CC ACT\_SITE 75 75  
 CC ACT\_SITE 121 121  
 CC ACT\_SITE 214 214  
 CC CARBOHYD 114 114  
 CC DISULFID 19 141  
 CC DISULFID 60 76  
 CC DISULFID 155 220  
 CC DISULFID 187 201  
 CC DISULFID 210 239  
 CC SEQUENCE 264 AA; 28002 MW; 3F629F02FA6DDB4 CRC64;

Query Match 72.1%; Score 49; DB 1; Length 264;  
 Best Local Similarity 72.7%; Pred. No. 0.35;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 CGVPAIQPVLS 11  
 |||||  
 DB 19 CGIPAIPKALS 29

RESULT 7  
 CTRA\_GADMO STANDARD; PRT; 263 AA.  
 AC P47796;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Chymotrypsin A precursor (EC 3.4.21.1).  
 OS Gadus morhua (Atlantic cod).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.  
 OX NCBI\_TaxID=8049;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pyloric caeca;  
 RX MEDLINE=94368860; PubMed=8086467;  
 RA Gudmundsdottir A., Oskarsson S., Eskin A.E., Craik C.S.,  
 RA Bjarnason J.B., Bjarnason J.B.;  
 RA "Atlantic cod cDNA encoding a psychrophilic chymotrypsinogen.";   
 RL Biochim. Biophys. Acta 1219:211-214 (1994).  
 RN [2]  
 RP SEQUENCE OF 19-30 AND 34-49.

CC TISSUE=Pyloric caeca;  
 CC MEDLINE=92111252; PubMed=1764912;  
 CC Asgeirsson B., Bjarnason J.B.;  
 CC "Structural and kinetic properties of chymotrypsin from Atlantic cod  
 CC (Gadus morhua). Comparison with bovine chymotrypsin.";   
 CC RT

RL Comp. Biochem. Physiol. 99B:327-335(1991).  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,  
 CC Phe-|-Xaa, Leu-|-Xaa  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).

CC EMBL; X78490; CAA55242.1; -  
 CC HSSP; P00766; LCHG.  
 CC MEROPS; S01.152; -  
 CC InterPro; IPR001314; Chymotrypsin.  
 CC InterPro; IPR001254; Ser. protease\_Try.  
 CC Pfam; PF00089; trypsin; 1.  
 CC PRINTS; PR00722; CHYMOTRYPSIN.  
 CC SMART; SM00200; Tryp\_Spc; 1.  
 CC PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
 CC Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal.  
 CC SIGNAL 1 18  
 CC CHAIN 19 263  
 CC ACT\_SITE 75 75  
 CC ACT\_SITE 120 120  
 CC ACT\_SITE 213 213  
 CC DISULFID 19 140  
 CC DISULFID 60 76  
 CC DISULFID 154 219  
 CC DISULFID 186 200  
 CC DISULFID 209 238  
 CC DISULFID 21 21  
 CC CONFLICT 25 25  
 CC CONFLICT 29 29  
 CC CONFLICT 44 44  
 CC CONFLICT 46 46  
 CC SEQUENCE 263 AA; 28294 MW; 47AAC699A0A64FBB CRC64;

Query Match 69.1%; Score 47; DB 1; Length 263;  
 Best Local Similarity 66.7%; Pred. No. 0.78;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 OY 1 CGVPAIQPVLSG 12  
 |||||  
 DB 19 CGRPAISPVTG 30

RESULT 8  
 CTRB\_GADMO STANDARD; PRT; 245 AA.  
 AC P80646;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Chymotrypsin B (EC 3.4.21.1).  
 OS Gadus morhua (Atlantic cod).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.  
 OX NCBI\_TaxID=8049;  
 RN [1]  
 RP SEQUENCE.

CC TISSUE=Pyloric caeca;  
 CC MEDLINE=96439045; PubMed=8841380;  
 CC Leth-Larsen R., Asgeirsson B., Thorolfsson M., Noerregaard-Madsen M.,  
 CC Hoejrup P.;  
 CC "Structure of chymotrypsin variant B from Atlantic cod, Gadus  
 CC morhua.";   
 CC RT

```

RN Blochim. Biophys. Acta 1297:49-56(1996).
RP SEQUENCE OF 1-12 AND 16-31.
RC TISSUE=Pyoric caeca; PubMed=1764912;
RX MEDLINE=92111252; PubMed=1764912;
RA Agelarsen B., Bjarnason J.B.;
RT "Structural and kinetic properties of chymotrypsin from Atlantic cod
  (Gadus morhua). Comparison with bovine chymotrypsin.";
RL Comp Biochem. Physiol. 99B:327-335(1991).
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
CC Phe-|-Xaa, Leu-|-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR HSP; P00766; 1CHG.
DR MEROPS; S01.152; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR PRINTS; PF00089; trypsin; 1.
DR SMART; SM00020; TRY_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS0134; TRYPSIN_HIS; 1.
DR PROSITE; PS0135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen.
FT CHAIN 1 13 CHYMOTRYPSIN B, A CHAIN.
FT CHAIN 16 245 CHYMOTRYPSIN B, B CHAIN.
FT ACT_SITE 57 57 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 101 101 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 195 195 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 1 121 BY SIMILARITY.
FT DISULFID 42 58 BY SIMILARITY.
FT DISULFID 135 201 BY SIMILARITY.
FT DISULFID 167 182 BY SIMILARITY.
FT DISULFID 191 220 BY SIMILARITY.
FT CONFLICT 9 11 QVT -> VIS (IN REF. 2).
FT CONFLICT 26 26 S -> T (IN REF. 2).
FT CONFLICT 28 29 PW -> Y (IN REF. 2).
SQ SEQUENCE 245 AA; 26260 MW; 74FE0D425517AB02 CRC64;

Query Match 67.6%; Score 46; DB 1; Length 245;
Best Local Similarity 66.7%; Pred. No. 1.1;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGVPAIQPVLSG 12
DB 1 CGSPAIPQVVTG 12
||| ||||| :||
||| ||||| :||

RESULT 9
RBL_CHLRE STANDARD; PRT; 475 AA.
ID_RBL_CHLRE AC P00877;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Ribulose biphosphate carboxylase large chain precursor (EC 4.1.1.39)
DE (RuBisCO large subunit).
GN RECL.
OS Chlamydomonas reinhardtii.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A. PubMed=6302265;
RX MEDLINE=83189072; PubMed=6302265;
RA Dron M., Rahire M., Rochaix J.-D.;
RT "Sequence of the chloroplast DNA region of Chlamydomonas reinhardtii
  containing the gene of the large subunit of ribulose biphosphate
  carboxylase and parts of its flanking genes.";
RL J. Mol. Biol. 162:775-793(1982).
RN [2]
RP SEQUENCE OF 3-14, AND ACETYLATION.

```

```

RA Houtz R.L., Poneleit L., Jones S.B., Royer M., Stults J.T.;
RT "Posttranslational modifications in the amino-terminal region of the
  large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from
  several plant species.";
RL Plant Physiol. 98:1170-1174(1992).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
  D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
  CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
  THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
  REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
  ACTIVE SITE.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-
  phospho-D-glycerate.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) =
  3-phospho-D-glycerate + 2-phosphoglycolate.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J01399; AAA84449.1; -.
CC PIR; A01097; RKKML.
CC InterPro; IPR000685; RuBisCO_large.
CC Pfam; PF00016; RuBisCO_large; 1.
CC Pfam; PF02788; RuBisCO_large; 1.
CC PROSITE; PS00157; RUBISCO_LARGE; 1.
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
  Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Acetylation.
FT PROPEP 1 2 RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE
FT CHAIN 3 475 CHAIN.
FT MOD_RES 3 3 ACETYLATION
FT ACT_SITE 201 201 BINDING OF CO(2) ACTIVATES THE ENZYME.
SQ SEQUENCE 475 AA; 52543 MW; 5A9BFD394CF7D4D4 CRC64;

Query Match 58.8%; Score 40; DB 1; Length 475;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 CGVPAIQPVLSG 12
DB 369 CSMPGVMPVASG 380
||| :|||
||| :|||

RESULT 10
HE_HEMPU STANDARD; PRT; 591 AA.
ID_HE_HEMPU AC P9793;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hatching enzyme precursor (EC 3.4.24.12) (HE) (HEZ) (Envelysin)
DE (Sea-urchin-hatching proteinase).
DE Hemacentrotus pulcherrimus (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
OC Hemacentrotus.
OX NCBI_TaxID=7650;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 170-220 AND 504-528.
RC TISSUE=Blasculata;
RX MEDLINE=97332490; PubMed=9188724;
RA Nomura K., Shimizu T., Kinoh H., Sendai Y., Inomata M., Suzuki N.;
RT "Sea urchin hatching enzyme (envelysin): cDNA cloning and deprivation
  of protein substrate specificity by autolytic degradation.";
RL Biochemistry 36:7225-7238(1997).

```

[2]  
 RP CLEAVAGE SPECIFICITY.  
 RX MEDLINE=91283448; PubMed=1711895;  
 RA Nomura K., Tanaka H., Kikkawa Y., Yamaguchi M., Suzuki N.;  
 RT "The specificity of sea urchin hatching enzyme (envelysin) places it  
 RL in the mammalian matrix metalloproteinase family.";  
 RN Biochemistry 30:6115-6123(1991).  
 [3]  
 RP STEREO-SPECIFICITY.  
 RX MEDLINE=93223852; PubMed=8467915;  
 RA Nomura K., Suzuki N.;  
 RT "Stereo-specific inhibition of sea urchin envelysin (hatching enzyme)  
 RL by a synthetic autoinhibitor peptide with a cysteine-switch consensus  
 RN sequence.";  
 FEBS Lett. 321:84-88(1993).  
 CC -!- FUNCTION: ALLOWS THE SEA URCHIN TO DIGEST THE PROTECTIVE ENVELOPE  
 CC DERIVED FROM THE EGG EXTRACELLULAR MATRIX THUS ALLOWING THE SEA  
 CC URCHIN TO SWIM FREELY.  
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage: on the amino side of  
 CC bulky hydrophobic residues, -Leu, -Ile, -Phe, as well as -Tyr.  
 CC -!- SUBUNIT: DURING HATCHING THE 50 kDa MATURE ENZYME IS  
 CC AUTOLYTIKALLY CLEAVED TO PRODUCE A MAJOR 38 kDa AND A MINOR 15 kDa  
 CC FORM WHICH MAY BE DISULFIDE LINKED. SUBSEQUENT CLEAVAGE OF THE 38  
 CC kDa SPECIES YIELDS A 32 kDa NON-SPECIFIC PROTEASE.  
 CC -!- DEVELOPMENTAL STAGE: EMBRYO, BLASTULA STAGE. HIGHEST ACTIVITY AT  
 CC 12.5 HRS EMBRYO STAGE.  
 CC -!- DOMAIN: THERE ARE TWO DISTINCT DOMAINS IN THIS PROTEIN; THE  
 CC CATALYTIC N-TERMINAL, AND THE C-TERMINAL WHICH IS INVOLVED IN  
 CC SUBSTRATE SPECIFICITY.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A.  
 CC -!- SIMILARITY: CONTAINS 1 HEMOPEXIN-LIKE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB000719; BAA19171.1; -;  
 DR HSSP; P03956; ICSL.  
 DR MEROPS; M10.010; -;  
 DR InterPro; IPR000585; Hemopexin.  
 DR InterPro; IPR001818; Matrxin.  
 DR InterPro; IPR000130; Zn\_Mtpeptase.  
 DR Pfam; PF00045; hemopexin; 4.  
 DR Pfam; PF00413; Peptidase M10; 1.  
 DR PRINTS; PR00138; MATRXIN.  
 DR SMART; SM00120; HX; 4.  
 DR SMART; SM00235; ZnMc; 1.  
 DR PROSITE; PS00024; HEMOPEXIN; 1.  
 DR PROSITE; PS00142; ZINC PROTEASE; 1.  
 DR PROSITE; PS00546; CYSTEINE SWITCH; 1.  
 KW Hydrolyase; Metalloprotease; Zinc; Glycoprotein; Zymogen;  
 KW Autocatalytic cleavage; Signal; Calcium.  
 FT SIGNAL 1 18 PROBABLE.  
 FT PROPEP 19 169 ACTIVATION PEPTIDE (BY SIMILARITY).  
 FT CHAIN 170 591 50 kDa HATCHING ENZYME.  
 FT CHAIN 170 450 32 kDa HATCHING ENZYME (NON-SPECIFIC).  
 FT CHAIN 170 503 38 kDa HATCHING ENZYME.  
 FT CHAIN 504 591 15 kDa PEPTIDE.  
 FT DOMAIN 62 76 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 336 380 ARG/THR-RICH (HINGE REGION).  
 FT DOMAIN 380 589 HEMOPEXIN-LIKE.  
 FT DOMAIN 448 451 POLY-TYR.  
 FT SITE 162 162 CYSTEINE SWITCH (POTENTIAL).  
 FT METAL 286 286 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT SITE 287 287 BY SIMILARITY.  
 FT METAL 290 290 ZINC (CATALYTIC).  
 FT SITE 450 451 CLEAVAGE (AUTOLYTIC DURING HATCHING).  
 FT SITE 450 451 CLEAVAGE (AUTOLYTIC DURING HATCHING).  
 FT DISULFID 383 586 BY SIMILARITY.

FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 588 588 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 195 195 S -> N (IN REF. 1; AA SEQUENCE).  
 FT CONFLICT 197 197 T -> N (IN REF. 1; AA SEQUENCE).  
 FT CONFLICT 203 203 L -> I (IN REF. 1; AA SEQUENCE).  
 FT CONFLICT 215 216 SL -> GN (IN REF. 1; AA SEQUENCE).  
 FT CONFLICT 220 220 E -> N (IN REF. 1; AA SEQUENCE).  
 FT CONFLICT 509 509 P -> R (IN REF. 1; AA SEQUENCE).  
 FT CONFLICT 511 512 SS -> RR (IN REF. 1; AA SEQUENCE).  
 FT CONFLICT 518 518 P -> L (IN REF. 1; AA SEQUENCE).  
 FT CONFLICT 522 522 R -> I (IN REF. 1; AA SEQUENCE).  
 SQ SEQUENCE 591 AA; 66126 MW; 5DCB448C6758C70D CRC64;  
 Query Match 58.8%; Score 40; DB 1; Length 591;  
 Best Local Similarity 58.3%; Pred. No. 27;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 CGVPAIQPVLSG 12  
 Db 162 CGVPIIPVITG 173  
 RESULT 11  
 Y084 CAEEL  
 ID Y084 CAEEL STANDARD; PRT; 364 AA.  
 AC P34620;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical 42.2 kDa protein ZK1236.4 in chromosome III.  
 GN ZK1236.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1] \_TaxID=6239;  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=Briscot N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,  
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 RA Johnston L., Jones M., Kersey J., Kirsten J., Laister N.,  
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
 RA Sims M., Smailson N., Smith A., Smith M., Sonhammer E., Staden R.,  
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
 RA Woldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 CC -!- SIMILARITY: TO TRANSPOSON TL-2 OF AFRICAN MALARIA MOSQUITO.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L13200; AAA28194.1; -;  
 DR PIR; S44899; S44899.  
 DR WormPep; ZK1236.4; CE00531.  
 DR InterPro; IPR000477; RVTse.  
 DR Pfam; PF00078; rvt; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 364 AA; 42163 MW; B7BACEF8FEDFB0B CRC64;

Wed Feb 12 11:59:41 2003

Query Match 58.1%; Score 39.5; DB 1; Length 364;  
 Best Local Similarity 56.2%; Pred. No. 20;  
 Matches 9; Conservative 1; Mismatches 3; Indels 3; Gaps 1;

OY 1 CGVP---AIQPVLSGL 13  
 ||||| |||||  
 Db 79 CGVPQGSVSPVFGI 94

## RESULT 12

ID \_PRLA\_LYSEN STANDARD; PRT; 397 AA.  
 AC P00778;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Alpha-lytic protease precursor (EC 3.4.21.12) (Alpha-lytic  
 DE endopeptidase).  
 GN ALPHA-LP.  
 OS Lysobacter entomogenes.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OC Lysobacter.  
 OX NCBI\_TaxID=69;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 29487;  
 RX MEDLINE=89034140; PubMed=3053694;  
 RA Epstein D.M., Wensink P.C.;  
 RT "The alpha-lytic protease gene of Lysobacter entomogenes. The  
 RT nucleotide sequence predicts a large prepro-peptide with homology to  
 RT pro-peptides of other chymotrypsin-like enzymes.";  
 RL J. Biol. Chem. 263:16586-16590(1988).  
 RN [2]  
 RP REVISIONS.  
 RA Epstein D.M.;  
 RL Submitted (MAY-1989) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 29487;  
 RX MEDLINE=89172068; PubMed=3234766;  
 RA Silen J.L., McGrath C.N., Smith K.R., Agard D.A.;  
 RT "Molecular analysis of the gene encoding alpha-lytic protease:  
 RT evidence for a preproenzyme.";  
 RL Gene 69:237-244(1988).  
 RN [4]  
 RP SEQUENCE OF 199-396.  
 RC STRAIN=ATCC 29487;  
 RX MEDLINE=71039222; PubMed=5482494;  
 RA Olson M.O.J., Nagabhushan N., Dzwiniel M., Smillie L.B.,  
 RA Whitaker D.R.;  
 RT "Primary structure of alpha-lytic protease: a bacterial homologue of  
 RT the pancreatic serine proteases.";  
 RL Nature 228:438-442(1970).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
 RX MEDLINE=80074995; PubMed=117110;  
 RA Brayer G.D., Delbaere L.T.J., James M.N.G.;  
 RT "Refined structure of the alpha-lytic protease from Myxobacter 495  
 RT at 2.8-A resolution.";  
 RL J. Mol. Biol. 131:743-775 (1979).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).  
 RX MEDLINE=86011557; PubMed=3900416;  
 RA Fujinaga M., Delbaere L.T.J., Brayer G.D., James M.N.G.;  
 RT "Refined structure of alpha-lytic protease at 1.7-A resolution.  
 RT Analysis of hydrogen bonding and solvent structure.";  
 RL J. Mol. Biol. 184:479-502(1985).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).  
 RC STRAIN=ATCC 29487;  
 RX MEDLINE=98393457; PubMed=9724517;  
 RA Peters R.J., Shiau A.K., Sohl J.L., Anderson D.E., Tang G.,  
 RA Silen J.L., Agard D.A.;

RT "Pro region C-terminus;Protease active site interactions are critical  
 RT in catalyzing the folding of alpha-lytic protease.";  
 RL Biochemistry 37:12058-12067(1998).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).  
 RX MEDLINE=99023190; PubMed=9808037;  
 RA Sauter N.K., Mau T., Rader S.D., Agard D.A.;  
 RT "Structure of alpha-lytic protease complexed with its pro region.";  
 RL Nat. Struct. Biol. 5:945-950(1998).  
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of proteins, especially bonds  
 CC adjacent to L-alanine and L-valine residues in bacterial cell  
 CC walls, elastin and other proteins  
 CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2A.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; J04052; AAA25409.1; --  
 CC EMBL; M22763; AAA74111.1; --  
 CC PIR; A31772; TRYXB4.  
 DR PDB; 2ALP; 15-OCT-89.  
 DR PDB; 1GBA; 29-JAN-96.  
 DR PDB; 1GBB; 29-JAN-96.  
 DR PDB; 1GBC; 29-JAN-96.  
 DR PDB; 1GBD; 29-JAN-96.  
 DR PDB; 1GBE; 29-JAN-96.  
 DR PDB; 1GBF; 29-JAN-96.  
 DR PDB; 1GBH; 29-JAN-96.  
 DR PDB; 1GBI; 29-JAN-96.  
 DR PDB; 1GBJ; 29-JAN-96.  
 DR PDB; 1GBK; 29-JAN-96.  
 DR PDB; 1GBL; 29-JAN-96.  
 DR PDB; 1GBM; 29-JAN-96.  
 DR PDB; 2LPR; 15-JAN-93.  
 DR PDB; 3LPR; 15-JAN-93.  
 DR PDB; 5LPR; 15-JAN-93.  
 DR PDB; 6LPR; 15-JAN-93.  
 DR PDB; 7LPR; 15-JAN-93.  
 DR PDB; 8LPR; 15-JAN-93.  
 DR PDB; 9LPR; 15-JAN-93.  
 DR PDB; 1P01; 15-APR-90.  
 DR PDB; 1P02; 15-APR-90.  
 DR PDB; 1P03; 15-APR-90.  
 DR PDB; 1P04; 15-APR-90.  
 DR PDB; 1P05; 15-APR-90.  
 DR PDB; 1P06; 15-APR-90.  
 DR PDB; 1P09; 15-APR-90.  
 DR PDB; 1P10; 15-APR-90.  
 DR PDB; 1P11; 15-JAN-93.  
 DR PDB; 1P12; 15-JAN-93.  
 DR PDB; 1TAL; 01-APR-97.  
 DR PDB; 2ULL; 07-JUL-97.  
 DR PDB; 1BQO; 12-AUG-98.  
 DR PDB; 1QQ4; 03-MAY-00.  
 DR PDB; 1QRW; 03-MAY-00.  
 DR PDB; 1QRX; 18-JUN-99.  
 DR PDB; 2PRO; 27-APR-99.  
 DR PDB; 3PRO; 27-APR-99.  
 DR PDB; 4PRO; 18-MAY-99.  
 DR MEROPS; S01.268; --  
 DR InterPro; IPR004236; AL protease.  
 DR InterPro; IPR001316; Endoptdse2A.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR Pfam; PF02983; AL protease; 2.  
 DR PRINTS; PR00861; ALYTICPTASE.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.



```

KW Hydrolase; Serine protease; 3D-structure; Zymogen; Signal.
FT SIGNAL 1 24
FT PROPEP 25 199
FT CHAIN 200 397
FT DISULFID 216 336
FT DISULFID 300 310
FT DISULFID 336 369
FT ACT_SITE 235 235
FT ACT_SITE 262 262
FT ACT_SITE 342 342
FT ACT_SITE 358 358
FT ACT_SITE 371 371
FT CONFLICT 171 171
FT STRAND 201 204
FT TURN 205 205
FT STRAND 207 210
FT TURN 211 213
FT STRAND 214 217
FT STRAND 220 224
FT TURN 225 226
FT STRAND 227 232
FT HELIX 234 236
FT TURN 239 240
FT STRAND 242 245
FT TURN 246 247
FT STRAND 248 257
FT STRAND 260 260
FT STRAND 263 268
FT TURN 270 271
FT STRAND 273 280
FT TURN 281 282
FT STRAND 283 286
FT STRAND 289 289
FT TURN 295 296
FT STRAND 298 303
FT TURN 304 306
FT STRAND 307 322
FT TURN 323 324
FT STRAND 325 333
FT STRAND 338 338
FT TURN 339 340
FT TURN 342 343
FT STRAND 345 347
FT TURN 349 350
FT STRAND 352 360
FT TURN 364 365
FT STRAND 367 367
FT TURN 369 370
FT HELIX 373 375
FT STRAND 378 382
FT HELIX 384 390
FT TURN 391 391
FT STRAND 393 394
SQ SEQUENCE 397 AA; 41077 MW; 267FB6BFBF57F33CB CRC64;

Query Match 58.1%; Score 39.5; DB 1; Length 397;
Best Local Similarity 40.0%; Pred. No. 22;
Matches 8; Conservative 3; Mismatches 0; Indels 9; Gaps 1;

QY 1 CGVPA-----IQPVLS 11
|||
Db 369 CGIPASQRSSIFERLQPLLS 368
|||

RESULT 13
ID_PPNK_HALN1 STANDARD; PRT; 282 AA.
AC QPHNX7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable inorganic polyphosphate/ATP-NAD kinase (EC 2.7.1.23)
DE (Poly(P)/ATP NAD kinase).
GN PPNK OR VNG1900C.

```

```

OS Halobacterium sp. (strain NRC-1)
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OX NCBI_TaxID=64051;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sirogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welter R., Goo Y.A.,
RA Lettshauser B., Keller K., Cruz R., Banson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Alam M., Freitas T., Peck R.F., Pohlischer M., Spudich J.L., Jung K.-H.,
RA Ehardt H., Lowe T.M., Liang P., Daniels C.J., Dennis P.P., Omer A.D.,
RT "Genome sequence of Halobacterium species NRC-1."
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181 (2000).
CC -|- FUNCTION: Catalyzes the phosphorylation of NAD to NADP. Utilizes
CC ATP and other nucleoside triphosphates as well as inorganic
CC polyphosphate as a source of phosphorus (By similarity).
CC -|- CATALYTIC ACTIVITY: ATP + NAD(+) = ADP + NADP(+).
CC -|- COFACTOR: Requires divalent metal ions for activity (By
CC similarity).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -|- SIMILARITY: BELONGS TO THE NAD KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE005089; AAG20093.1; -.
CC InterPro; IPR002504; ATP NADK.
CC Pfam; PF01513; NAD kinase; 1.
CC Transferrase; Kinase; NAD; NADP; Complete proteome.
SQ SEQUENCE 282 AA; 29120 MW; 5DA1AF25147D7500 CRC64;

Query Match 57.4%; Score 39; DB 1; Length 282;
Best Local Similarity 46.2%; Pred. No. 19;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLSGL 13
|||
Db 207 CGADALPPLVTGL 219
|||

RESULT 14
ID_UL21_PRVN3 STANDARD; PRT; 523 AA.
AC Q00703;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE Protein UL21 homolog.
OS Pseudorabies virus (strain NIA-3) (PRV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10349;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93059656; PubMed=1331512;
RA Berns A.A., Kimman T.T., Pol J.J., Wagenaar F.F., de Wind N.N.;
RT "The pseudorabies virus homology of the herpes simplex virus UL21
RT gene product is a capsid protein which is involved in capsid
RT maturation."
RL J. Virol. 66:7096-7103 (1992).
CC -|- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL21,
CC HSV-1 40, EBV-4 UL21, AND VZV 38.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M95285; AAA47474.1; --  
CC PIR; A44195; A44195.  
CC InterPro; IPR004936; Herpes\_UL21.  
CC Pfam; PF03252; UL21; 1.  
CC SEQUENCE 523 AA; 55019 MW; 9E69949DAE94CAC6 CRC64;  
CC

Query Match 57.48; Score 39; DB 1; Length 523;  
Best Local Similarity 54.58; Pred. No. 35;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GVPATQPVLSG 12  
DB 187 GIPGVRPLSG 197

RESULT 15  
NUOG SALTY STANDARD; PRT; 907 AA.  
AC P33900;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE NADH dehydrogenase I chain G (EC 1.6.5.3) (NADH-ubiquinone  
oxidoreductase chain G) (NUO7)  
GN NUOG OR STM2323 OR STY2553  
OS Salmonella typhimurium, and  
OS Salmonella typhi  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=602, 601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollk S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium  
LT2."; Nature 413:852-856 (2001).  
RL [2]  
RN SEQUENCE OF 1-611 FROM N.A.  
RP SPECIES=S.typhimurium;  
RC MEDLINE=94052195; PubMed=8234329;  
RX Archer C.D., Wang X., Elliott T.;  
RA "Mutants defective in the energy-conserving NADH dehydrogenase of  
Salmonella typhimurium identified by a decrease in energy-dependent  
proteolysis after carbon starvation."; Proc. Natl. Acad. Sci. U.S.A. 90:9877-9881 (1993).  
RL [3]  
RN SEQUENCE OF 611-907 FROM N.A.  
RP SPECIES=S.typhimurium; STRAIN=P98;  
RX MEDLINE=98037521; PubMed=931470;  
RA Zhang-Barber L.Z., Turner A.K., Martin G., Fraenkel G., Dougan G.,  
RA Barrow P.A.;  
RA "Influence of genes encoding proton-translocating enzymes on  
suppression of Salmonella typhimurium growth and colonization."; J.  
Bacteriol. 179:7186-7190 (1997).  
RL [4]  
RN SEQUENCE FROM N.A.  
RP SPECIES=S.typhi; STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parkhill J., Dougan G., James K.B., Thomson N.R., Pickard D., Wain J.,  
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,

BA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,  
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagals K.,  
RA Krogg A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
RA Whitehead S., Barrett B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmonella  
enterica serovar Typhi CT18";  
RL Nature 413:848-852 (2001).  
CC -!- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY  
CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED  
CC TO BE UBIQUINONE. DOES COUPLE THE REDOX REACTION TO PROTON  
CC TRANSLOCATION AND THUS CONSERVES THE REDOX ENERGY IN A PROTON  
CC GRADIENT.  
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
CC -!- COFACTOR: BINDS 1 2FE-2S CLUSTER AND 1 4FE-4S CLUSTER (POTENTIAL).  
CC -!- SUBUNIT: COMPOSED OF 13 DIFFERENT SUBUNITS. SUBUNITS NUOCD, E,  
CC F, AND, G CONSTITUTE THE PERIPHERAL SECTOR OF THE COMPLEX.  
CC -!- SIMILARITY: BELONGS TO THE COMPLEX I 75 Kda SUBUNIT FAMILY.  
CC -!- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO  
CC FRAMESHIFTS IN POSITIONS 714 AND 805.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AB008804; AAL21224.1; ALT\_INIT.  
DR EMBL; L22504; AAA16063.1; ALT\_INIT.  
DR EMBL; L42521; -; NOT ANNOTATED CDS.  
DR EMBL; AL627274; CAD07555.1; ALT\_INIT.  
DR StyGene; SG10257; nuog.  
DR InterPro; IPR000283; Complex1\_75K.  
DR InterPro; IPR001041; Ferredoxin.  
DR InterPro; IPR001467; Prok\_Mboxred.  
DR Pfam; PF00111; fer2; 1.  
DR Pfam; PF00384; molybdopterin; 1.  
DR PROSITE; PS00641; COMPLEX1\_75K\_1; 1.  
DR PROSITE; PS00642; COMPLEX1\_75K\_2; 1.  
DR PROSITE; PS00643; COMPLEX1\_75K\_3; 1.  
KW Oxidoreductase; NAD; Ubiquinone; Iron-sulfur; 4Fe-4S;  
KW Complete proteome.  
FT INIT MET 0 BY SIMILARITY.  
FT METAL 22 22 IRON-SULFUR (2FE-2S) (POTENTIAL).  
FT METAL 33 33 IRON-SULFUR (2FE-2S) (POTENTIAL).  
FT METAL 44 44 IRON-SULFUR (2FE-2S) (POTENTIAL).  
FT METAL 47 47 IRON-SULFUR (2FE-2S) (POTENTIAL).  
FT METAL 102 102 IRON-SULFUR (2FE-2S) (POTENTIAL).  
FT METAL 105 105 IRON-SULFUR (2FE-2S) (POTENTIAL).  
FT METAL 111 111 IRON-SULFUR (2FE-2S) (POTENTIAL).  
FT METAL 150 150 IRON-SULFUR (4FE-4S) (POTENTIAL).  
FT METAL 153 153 IRON-SULFUR (4FE-4S) (POTENTIAL).  
FT METAL 156 156 IRON-SULFUR (4FE-4S) (POTENTIAL).  
FT METAL 200 200 IRON-SULFUR (4FE-4S) (POTENTIAL).  
FT METAL 677 677 A -> V (IN REF. 3).  
FT CONFLICT 785 785 I -> L (IN REF. 3).  
FT CONFLICT 846 846 A -> T (IN REF. 3).  
FT CONFLICT 853 853 A -> T (IN REF. 3).  
FT CONFLICT 887 887 M -> I (IN REF. 3).  
SQ SEQUENCE 907 AA; 99896 MW; 5DE802178260C65E CRC64;

Query Match 57.48; Score 39; DB 1; Length 907;  
Best Local Similarity 63.64; Pred. No. 59;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GVPATQPVLSG 12  
DB 886 GMPGIAPVLG 896

Wed Feb 12 11:59:41 2003

us-10-036-371-5.rsp

Page 11

Search completed: February 12, 2003, 10:23:17  
Job time : 3.74627 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:16:06 ; Search time 9.19701 Seconds  
(without alignments)  
291.248 Million cell updates/sec

Title: US-10-036-371-5

Perfect score: 68

Sequence: 1 CGVPAIQPVLSGL 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_muc.\*

8: sp\_organelle.\*

9: sp\_phage.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_virus.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	65	95.6	263	11 Q9DC86	Q9dc86 mus musculus
2	65	95.6	263	11 Q9D8X8	Q9d8x8 mus musculus
3	65	95.6	263	11 Q9CR35	Q9cr35 mus musculus
4	54	79.6	18	4 Q9UCF4	Q9ucf4 homo sapien
5	53	77.9	25	6 Q9RL4	Q9trl4 sus scrofa
6	48	70.6	24	13 Q9RL51	Q9prl1 gadus morhu
7	48	70.6	261	13 Q9W704	Q9w7q4 paralichthy
8	48	70.6	264	11 Q9D960	Q9d960 mus musculus
9	48	70.6	264	11 Q9D7P8	Q9d7p8 mus musculus
10	48	70.6	264	11 Q9E028	Q9ezt8 rattus norv
11	48	70.6	264	11 Q9ER05	Q9er05 mus musculus
12	47	69.1	23	13 Q9PS0	Q9ps0 gadus morhu
13	47	69.1	260	13 Q9WQ3	Q9wq3 paralichthy
14	46	67.6	263	13 Q9PW06	Q9pw06 gadus morhu
15	46	67.6	376	8 Q9XR04	Q9xr04 chlamydomon
16	46	67.6	376	8 Q9XR03	Q9xr03 chlamydomon

17	46	67.6	376	8 Q9XR02	Q9xr02 chlamydomon
18	46	67.6	376	10 Q65774	Q65774 chlamydomon
19	46	67.6	376	10 Q65777	Q65777 chlamydomon
20	45	66.2	519	2 Q9F5J5	Q9f5j5 streptomyce
21	44	65.4	384	16 Q8ZF12	Q8zf12 yersinia pe
22	44	64.7	40	13 Q9PSF2	Q9psp2 gallus gall
23	44	64.7	205	4 Q96FA4	Q96fa4 homo sapien
24	44	64.7	270	5 Q96871	Q96871 trichinella
25	43	63.2	391	2 Q88172	Q88172 escherichia
26	42	61.8	376	8 Q98298	Q98298 chlamydomon
27	41	60.3	342	16 Q8YD6	Q8yfd6 brucella me
28	40	58.8	153	8 Q95H37	Q95h37 eudorina ml
29	40	58.8	172	10 Q8W2X4	Q8w2x4 oryza sativ
30	40	58.8	195	10 Q946P4	Q946p4 uncultured
31	40	58.8	255	16 Q9A4N6	Q9a4n6 caulobacter
32	40	58.8	326	16 Q82MQ2	Q82mq2 salmoneilla
33	40	58.8	376	8 Q24619	Q24619 platydorina
34	40	58.8	376	8 Q24628	Q24628 pandorina u
35	40	58.8	376	8 Q24636	Q24636 eudorina el
36	40	58.8	376	8 Q24652	Q24652 eudorina u
37	40	58.8	376	8 Q24655	Q24655 pandorina u
38	40	58.8	376	8 Q9ZXT3	Q9zxt3 chlamydomon
39	40	58.8	376	8 Q9T3M7	Q9t3m7 chlamydomon
40	40	58.8	376	8 Q9T3J9	Q9t3j9 chlamydomon
41	40	58.8	376	8 Q9G149	Q9g149 volvulina s
42	40	58.8	376	8 Q78227	Q78227 ymagishiel
43	40	58.8	376	8 Q9TH11	Q9th11 ymagishiel
44	40	58.8	376	8 Q98297	Q98297 chlamydomon
45	40	58.8	376	8 Q98300	Q98300 chlamydomon

#### ALIGNMENTS

RESULT 1  
Q9DC86 PRELIMINARY; PRT; 263 AA.  
ID Q9DC86 AC Q9DC86;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE 2200008D09RIK protein.  
GN 2200008D09RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]\_TaxID=10090;  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=SPLEEN;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,  
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wyrshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.

DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease.  
SQ SEQUENCE 263 AA; 27898 MW; C0638F98F905A92F CRC64;

Query Match 95.6%; Score 65; DB 11; Length 263;  
Best Local Similarity 92.3%; Pred. No. 0.0025; Gaps 0  
Matches 12; Conservative 1; Mismatches 0; Indels 0

QY 1 CGVPAIQPVLSGL 13  
|||||  
DB 19 CGVPAIQPVLTGL 31  
|||||

RESULT 3  
Q9CR35 PRELIMINARY; PRT; 263 AA.  
ID Q9CR35  
AC Q9CR35;  
DT 01-JUN-2001 (TRENBLrel. 17, Created)  
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
DE 220008D09RIK protein.  
DE 220008D09RIK.  
GN 220008D09RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OC NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=STOMACH, SPLEEN, AND PANCREAS;  
RX MEDLINE=21085560; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadoro K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Mateau Y., Nikaldo I., Pasole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bash G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.P.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wyszah-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtauki S.,  
RA Hayashizaki Y.,  
RA "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
DR ENBL; AK008927; BAB25971.1; -  
DR ENBL; AK003079; BAB22553.1; -  
DR ENBL; AK007765; BAB25241.1; -  
DR ENBL; AK007815; BAB25280.1; -  
DR ENBL; AK008729; BAB25861.1; -  
DR ENBL; AK008888; BAB25954.1; -  
DR HSP; P00766; 1GCT.  
DR MROPS; S01.152; -  
DR MGD; MGI:1913723; 220008D09RIK.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_Spc; 1.  
DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease.  
SQ SEQUENCE 263 AA; 27822 MW; 28C4487AFA1A26B27 CRC64;

Query Match 95.6%; Score 65; DB 11; Length 263;  
Best Local Similarity 92.3%; Pred. No. 0.0025;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLSGL 13  
|||||  
Db 19 CGVPAIQPVLTGL 31

## RESULT 4

Q9UCF4 ID Q9UCF4 PRELIMINARY; PRT; 18 AA.  
AC Q9UCF4;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE Chymotrypsinogen homolog (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=93285747; PubMed=8509158;  
RA Kim Y.J., Zhou Z., Hurtado J., Wood D.L., Choi A.S., Pescovitz M.D.,  
RA Warfel K.A., Vandagriff J., Davis J.K., Kwon B.S.;  
RT "IDDM patients' sera recognize a novel 30-kD pancreatic autoantigen  
related to chymotrypsinogen.";  
RL Immunol. Invest. 22:219-227(1993).  
SQ SEQUENCE 18 AA; 1826 MW; 38105D1037CA6837 CRC64;

Query Match 79.4%; Score 54; DB 4; Length 18;  
Best Local Similarity 91.7%; Pred. No. 0.014;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GVPAlQPVL SGL 13  
|||||  
Db 2 GVPAlHPVL SGL 13

## RESULT 5

Q9TRL4 ID Q9TRL4 PRELIMINARY; PRT; 25 AA.  
AC Q9TRL4;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Chymotrypsinogen homolog (Fragment).  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=93016107; PubMed=1400470;  
RA Gorr S.U., Hamilton J.W., Cohn D.V.;  
RT "Regulated, but not constitutive, secretory proteins bind porcine  
chymotrypsinogen.";  
RL J. Biol. Chem. 267:21595-21600(1992).  
DR HSSP; P00766; IACB.  
SQ SEQUENCE 25 AA; 2439 MW; F69AE003F1ACB136 CRC64;

Query Match 77.9%; Score 53; DB 6; Length 25;  
Best Local Similarity 91.7%; Pred. No. 0.03;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GVPAlQPVL SGL 13  
|||||  
Db 2 GVPAlPPVL SGL 13

## RESULT 6

Q9PRSI

ID Q9PRSI PRELIMINARY; PRT; 24 AA.  
AC Q9PRSI;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Chymotrypsin isoenzyme CHT1 (Fragment).  
OS Gadus morhua (Atlantic cod).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.  
OX NCBI\_TaxID=8049;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=96043258; PubMed=7584866;  
RA Raai A.J., Flengers R., Sletten K.;  
RT "Chymotrypsin isoenzymes in Atlantic cod; differences in kinetics and  
substrate specificity.";  
RL Comp. Biochem. Physiol. 112B:393-398(1995).  
DR HSSP; P00766; IACB.  
FT NON\_TER 1  
FT NON\_TER 24  
SQ SEQUENCE 24 AA; 2400 MW; 0A416ACA7B67F68D CRC64;

Query Match 70.6%; Score 48; DB 13; Length 24;  
Best Local Similarity 75.0%; Pred. No. 0.21;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLSG 12  
|||||  
Db 1 CGVPAIQPVSTG 12

## RESULT 7

Q9W7Q4 ID Q9W7Q4 PRELIMINARY; PRT; 261 AA.  
AC Q9W7Q4;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Chymotrypsinogen 1.  
OS Paralichthys olivaceus (Flounder).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;  
OC Pleuronectoidae; Paralichthyidae; Paralichthys.  
OX NCBI\_TaxID=8255;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PANCREAS;  
RA Suzuki T., Srivastava A.S., Kurokawa T.;  
RT "Japanese flounder mRNA for chymotrypsinogen 1.";  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.  
DR EMBL; AB029753; BAA82365.1; .  
DR HSSP; P00763; IDPO.  
DR MEROPS; S01.256; .  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; serine protease.  
SQ SEQUENCE 261 AA; 28184 MW; D7090A9D65395B7D CRC64;

Query Match 70.6%; Score 48; DB 13; Length 261;  
Best Local Similarity 66.7%; Pred. No. 2.2;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLSG 12

```

||||:|:|:|
17 CGVPSIKQVSG 28

Db 17 CGVPSIKQVSG 28

RESULT 8
Q9D960
ID Q9D960 PRELIMINARY; PRT; 264 AA.
AC Q9D960
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 1810004D15RIK protein.
GN CTRL OR 1810004D15RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.F.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlschki S.,
RA Hayaehizaki Y.;
RT "Functional annotation of a full-length mouse cdna collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AK007333; BAB24967.1; -.
DR HSSP; P00766; 4CHA.
DR MEROPS; S01.256; -.
DR MGD; MGI:88558; Ctrl.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 264 AA; 28151 MW; 1D979719E07C16DE CRC64;

Query Match 70.6%; Score 48; DB 11; Length 264;
Best Local Similarity 81.8%; Pred. No. 2.2;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLS 11
Db 19 CGVPAITPALS 29

RESULT 9
Q9D7P8
ID Q9D7P8 PRELIMINARY; PRT; 264 AA.
AC Q9D7P8
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 1810004D15RIK protein.
GN CTRL OR 1810004D15RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.F.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlschki S.,
RA Hayaehizaki Y.;
RT "Functional annotation of a full-length mouse cdna collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AK007333; BAB24967.1; -.
DR HSSP; P00766; 4CHA.
DR MEROPS; S01.256; -.
DR MGD; MGI:88558; Ctrl.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 264 AA; 28151 MW; 1D979719E07C16DE CRC64;

Query Match 70.6%; Score 48; DB 11; Length 264;
Best Local Similarity 81.8%; Pred. No. 2.2;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLS 11
Db 19 CGVPAITPALS 29

RESULT 9
Q9D7P8
ID Q9D7P8 PRELIMINARY; PRT; 264 AA.
AC Q9D7P8
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 1810004D15RIK protein.
GN CTRL OR 1810004D15RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.F.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlschki S.,
RA Hayaehizaki Y.;
RT "Functional annotation of a full-length mouse cdna collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AK007333; BAB24967.1; -.
DR HSSP; P00766; 4CHA.
DR MEROPS; S01.256; -.
DR MGD; MGI:88558; Ctrl.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 264 AA; 28151 MW; 1D979719E07C16DE CRC64;

Query Match 70.6%; Score 48; DB 11; Length 264;
Best Local Similarity 81.8%; Pred. No. 2.2;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLS 11
Db 19 CGVPAITPALS 29

RESULT 10
Q9EQZ8
ID Q9EQZ8 PRELIMINARY; PRT; 264 AA.
AC Q9EQZ8
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Chymopasin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RAT PANCREAS;
RX Sogame Y., Mitsu S., Kataoka K., Kato M., Sakagami J.,

```



RA Yamaguchi N.;  
 RT "Molecular cloning of rat chymopasins.";  
 CC Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 DR EMBL; AB020757; BAB20287.1; -.  
 DR HSSP; P00766; 4CHA.  
 DR MEROPS; S01.256; -.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; Tryp\_SPC; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR Hydrolase; Serine protease.  
 KW Hydrolyase; Serine protease.  
 SQ SEQUENCE 264 AA; 28116 MW; F9ED5D210FD3500E CRC64;

Query Match 70.6%; Score 48; DB 11; Length 264;  
 Best Local Similarity 81.8%; Pred. No. 2.2;  
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGVPAIQPVLS 11  
 Db 19 CGVPAITPALS 29  
 |||||  
 |||||

RESULT 11  
 Q9ER05 PRELIMINARY; PRT; 264 AA.  
 AC Q9ER05;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Chymopasin (Chymotrypsin A CTRA-1).  
 GN CTRL OR CTRA1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP Mitsui S., Yamaguchi N.;  
 RA "Molecular cloning of mouse chymopasin.";  
 RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 DR EMBL; AB016228; BAB20275.1; -.  
 DR EMBL; AF236365; AAL1034.1; -.  
 DR HSSP; P00766; 4CHA.  
 DR MEROPS; S01.256; -.  
 DR MCD; MGI:88558; Ctrl.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; Tryp\_SPC; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR Hydrolase; Serine protease.  
 KW Hydrolyase; Serine protease.  
 SQ SEQUENCE 264 AA; 28135 MW; 1D97970A07056C2 CRC64;

Query Match 70.6%; Score 48; DB 11; Length 264;  
 Best Local Similarity 81.8%; Pred. No. 2.2;  
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGVPAIQPVLS 11  
 Db 19 CGVPAITPALS 29  
 |||||  
 |||||

RESULT 12  
 Q9PR50 PRELIMINARY; PRT; 23 AA.  
 AC Q9PR50;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Chymotrypsin isoenzyme CHT2 (Fragment).  
 OS Gadus morhua (Atlantic cod).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.  
 OX NCBI\_TaxID=8049;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=96043258; PubMed=7584866;  
 RA Raad A.J., Flengsrud R., Sletten K.;  
 RT "Chymotrypsin isoenzymes in Atlantic cod; differences in kinetics and  
 RT substrate specificity.";  
 RL Comp. Biochem. Physiol. 112B:393-398 (1995).  
 FT NON\_TER 1 1  
 FT NON\_TER 23 23  
 SQ SEQUENCE 23 AA; 2356 MW; A5EC2A7B73EB3F5E CRC64;

Query Match 69.1%; Score 47; DB 13; Length 23;  
 Best Local Similarity 66.7%; Pred. No. 0.3;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGVPAIQPVLS 12  
 Db 1 CGRPAISPVTG 12  
 |||||  
 |||||

RESULT 13  
 Q9W7Q3 PRELIMINARY; PRT; 260 AA.  
 AC Q9W7Q3;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Chymotrypsinogen 2.  
 OS Paralichthys olivaceus (Flounder).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;  
 OC Pleuronectoidae; Paralichthyidae; Paralichthys.  
 OX NCBI\_TaxID=8255;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PANCREAS;  
 RA Suzuki T., Srivastava A.S., Kurokawa T.;  
 RT "Japanese flounder mRNA for chymotrypsinogen 2.";  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 DR EMBL; AB029754; BAB82366.1; -.  
 DR HSSP; P00766; 1CHG.  
 DR MEROPS; S01.152; -.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; Tryp\_SPC; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Serine protease.  
 SQ SEQUENCE 260 AA; 27793 MW; 9F583044E22F78C0 CRC64;

Query Match 69.1%; Score 47; DB 13; Length 260;  
 Best Local Similarity 66.7%; Pred. No. 3.3;  
 Matches 8; Conservative 2; Mismatches 0; Gaps 0;

QY 1 CGVPAIQPVLSG 12  
 |||||  
 DB 16 CGSPAIPQVTG 27

## RESULT 14

Q9PWQ6 PRELIMINARY; PRT; 263 AA.  
 AC Q9PWQ6;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Chymotrypsin B precursor (EC 3.4.21.1).  
 GN CHYB.  
 OS Gadus morhua (Atlantic cod).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.  
 OX NCBI\_TaxID=8049;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PYLORIC CAECA;  
 RX MEDLINE=20464334; PubMed=11011764;  
 RA Spilllaert R., Gudmundsdottir A.;  
 RT "Molecular Cloning of the Atlantic Cod Chymotrypsinogen B.";  
 RL Microb. Comp. Genomics 5:41-50(2000).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC EMBL; AJ242521; CAB43766.1; -.  
 DR HSSP; P00766; 1CHG.  
 DR MEROPS; S01.152; -.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS02040; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Serine protease; Signal.  
 FT SIGNAL 1 16 POTENTIAL.  
 FT CHAIN 17 263 CHYMOTRYPSIN B.  
 SQ SEQUENCE 263 AA; 28175 MW; EF61B18A34EE5E7C CRC64;

Query Match 67.6%; Score 46; DB 13; Length 263;  
 Best Local Similarity 66.7%; Pred. No. 5;  
 Matches 8; Conservative 2; Mismatches 0; Gaps 0;

QY 1 CGVPAIQPVLSG 12  
 |||||  
 DB 17 CGSPAIPQVTG 28

## RESULT 15

Q9XR04 PRELIMINARY; PRT; 376 AA.  
 AC Q9XR04;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RuBisCO  
 DE large subunit) (Fragment).  
 GN RBCL.  
 OS Chlamydomonas mutabilis.  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Chlamydomonadaceae; Chlamydomonas.  
 OX NCBI\_TaxID=47905;

RP SEQUENCE FROM N.A.  
 RC STRAIN=UTEX 578;  
 RA Morita E., Abe T., Tazuki M., Fujiwara S., Sato N., Hirata A.,  
 RA Sonoike K., Nozaki H.;  
 RT "Role of pyrenoids in the CO2-concentrating mechanism: comparative  
 RT morphology, physiology and molecular phylogenetic analysis of closely  
 RT related strains of Chlamydomonas and Chloromonas (Volvocales).";  
 RL Planta 208:365-372(1999).  
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-  
 CC RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
 CC ACTIVE SITE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-  
 CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-  
 CC PHOSPHO-D-GLYCERATE.  
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
 DR EMBL; AB022224; BAA82054.1; -.  
 DR InterPro; IPR000685; RuBisCO\_large.  
 DR Pfam; PF00016; RuBisCO\_large; 1.  
 DR Pfam; PF02788; RuBisCO\_large; 1.  
 DR PROSITE; PS00157; RUBISCO\_LARGE; 1.  
 KW Carbon dioxide fixation; Chloroplast; Lyase; Monooxygenase;  
 KW Oxidoreductase; Photorespiration; Photosynthesis.  
 FT NON\_TER 1 1  
 FT NON\_TER 376 376  
 SQ SEQUENCE 376 AA; 41774 MW; 845E743C36FB86A0 CRC64;

Query Match 67.6%; Score 46; DB 8; Length 376;  
 Best Local Similarity 58.3%; Pred. No. 7.1;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLSG 12  
 |||||  
 DB 359 CGLPGWMPVNSG 370

Search completed: February 12, 2003, 10:27:25  
 Job time : 11.197 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:18:40 ; Search time 2.79403 Seconds  
(without alignments)  
136.898 Million cell updates/sec

Title: US-10-036-371-5

Perfect score: 68

Sequence: 1 CGVPAIQPVLSGL 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*

1: /cgn2\_6/prodata/1/iaa/5A.COMB.pep:\*

2: /cgn2\_6/prodata/1/iaa/5B.COMB.pep:\*

3: /cgn2\_6/prodata/1/iaa/6A.COMB.pep:\*

4: /cgn2\_6/prodata/1/iaa/6B.COMB.pep:\*

5: /cgn2\_6/prodata/1/iaa/PCTUS.COMB.pep:\*

6: /cgn2\_6/prodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	61.8	326	US-08-988-876-1	Sequence 1, Appli
2	39.5	58.1	198	US-08-278-091-16	Sequence 16, Appl
3	39.5	58.1	198	US-08-483-859-16	Sequence 16, Appl
4	39.5	58.1	198	US-08-472-173-16	Sequence 16, Appl
5	39.5	58.1	198	US-08-487-167-16	Sequence 16, Appl
6	39.5	58.1	198	US-08-482-816-16	Sequence 16, Appl
7	39.5	58.1	198	US-08-296-149-16	Sequence 16, Appl
8	39.5	58.1	198	US-08-801-499-16	Sequence 16, Appl
9	39.5	58.1	198	US-08-615-271-16	Sequence 16, Appl
10	39.5	58.1	198	US-08-074-659-16	Sequence 16, Appl
11	39.5	58.1	198	US-09-106-468-16	Sequence 16, Appl
12	39.5	58.1	198	US-09-106-468A-16	Sequence 16, Appl
13	39.5	58.1	198	US-09-106-468A-16	Sequence 16, Appl
14	39.5	58.1	198	US-09-106-467-16	Sequence 16, Appl
15	39	57.4	525	US-08-681-129-2	Sequence 2, Appli
16	38	55.9	268	US-08-270-584A-2	Sequence 2, Appli
17	38	55.9	268	US-08-765-192-2	Sequence 2, Appli
18	38	55.9	268	US-09-199-793-2	Sequence 2, Appli
19	38	55.9	1048	US-09-171-699-10	Sequence 10, Appl
20	37	54.4	2441	US-08-194-468-2	Sequence 2, Appli
21	37	54.4	2441	US-08-961-739-2	Sequence 2, Appli
22	37	54.4	2441	US-09-514-247A-8	Sequence 8, Appli
23	36	52.9	70	US-08-935-450-12	Sequence 12, Appl
24	36	52.9	146	US-08-914-375C-37	Sequence 37, Appl
25	36	52.9	739	US-08-510-646B-33	Sequence 33, Appl
26	36	52.9	837	US-08-474-067-7	Sequence 7, Appli
27	36	52.9	837	US-08-474-068A-7	Sequence 7, Appli

Sequence 6, Appli  
Sequence 163, App  
Sequence 317, App  
Sequence 3728, Ap  
Sequence 5, Appli  
Sequence 4, Appli  
Sequence 3487, Ap  
Sequence 11, Appl  
Sequence 11, Appl  
Patent No. 5196194  
Sequence 387, App  
Sequence 2, Appli  
Sequence 8, Appli  
Sequence 18, Appl  
Sequence 18, Appl  
Sequence 3, Appli  
Sequence 127, App  
Sequence 13, Appl

ALIGNMENTS

RESULT 1  
US-08-988-876-1  
; Sequence 1, Application US/08988876  
; Patent No. 6063596  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Bandman, Olga  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Yue, Henry  
; TITLE OF INVENTION: G PROTEIN COUPLED RECEPTORS ASSOCIATED  
; TITLE OF INVENTION: WITH IMMUNE RESPONSE  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/988,876  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0441 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 326 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: PROSNOT01  
; CLONE: 364702  
; US-08-988-876-1

```
Query Match 61.8%; Score 42; DB 3; Length 326;
Best Local Similarity 66.7%; Pred. No. 15;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 GVPALQPVLSGL 13
DB 42 GDPELQPVLAGL 53

RESULT 2
US-08-278-091-16
; Sequence 16, Application US/08278091
; Patent No. 5506139
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/278,091
; FILING DATE: 21-JUL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-371
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 16:
; LENGTH: 198 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-278-091-16

Query Match 58.1%; Score 39.5; DB 1; Length 198;
Best Local Similarity 40.0%; Pred. No. 23;
Matches 8; Conservative 3; Mismatches 0; Indels 9; Gaps 1;

QY 1 CGVPA-----IQPVLS 11
DB 170 CGIPASQSSLFERLQFILLS 189

RESULT 3
US-08-483-859-16
; Sequence 16, Application US/08483859
; Patent No. 5656436
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/278,091
; FILING DATE: 21-JUL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-371
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 16:
; LENGTH: 198 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-278-091-16

Query Match 58.1%; Score 39.5; DB 1; Length 198;
Best Local Similarity 40.0%; Pred. No. 23;
Matches 8; Conservative 3; Mismatches 0; Indels 9; Gaps 1;

QY 1 CGVPA-----IQPVLS 11
DB 170 CGIPASQSSLFERLQFILLS 189

RESULT 4
US-08-472-173-16
; Sequence 16, Application US/08472173
; Patent No. 5665353
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/472,173  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/296,149  
; FILING DATE: 26-AUG-1994  
; APPLICATION NUMBER: US 08/278,091  
; FILING DATE: 21-JUL-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-493 MIS:vg  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 198 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-472-173-16

Query Match 58.1%; Score 39.5; DB 1; Length 198;  
Best Local Similarity 40.0%; Pred. No. 23;  
Matches 8; Conservative 3; Mismatches 0; Indels 9; Gaps 1;

QY 1 CGVPA-----IQPVLS 11  
|||:|  
Db 170 CGIPASQSSSLFERLPILS 189

RESULT 5  
US-08-487-167-16  
; Sequence 16, Application US/08487167  
; Patent No. 5869102  
; GENERAL INFORMATION:  
; APPLICANT: LOOSMORE, Sheena M.  
; APPLICANT: YANG, Yan-Ping  
; APPLICANT: CHONG, Pele  
; APPLICANT: OOMEN, Raymond P.  
; APPLICANT: KLEIN, Michel H.  
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with  
; TITLE OF INVENTION: Reduced Protease Activity  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.125  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/487,167  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/296,149  
; FILING DATE: 26-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-508 MIS:vg

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 198 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-487-167-16

Query Match 58.1%; Score 39.5; DB 2; Length 198;  
Best Local Similarity 40.0%; Pred. No. 23;  
Matches 8; Conservative 3; Mismatches 0; Indels 9; Gaps 1;

QY 1 CGVPA-----IQPVLS 11  
|||:|  
Db 170 CGIPASQSSSLFERLPILS 189

RESULT 6  
US-08-482-816-16  
; Sequence 16, Application US/08482816  
; Patent No. 5935573  
; GENERAL INFORMATION:  
; APPLICANT: LOOSMORE, Sheena M.  
; APPLICANT: YANG, Yan-Ping  
; APPLICANT: CHONG, Pele  
; APPLICANT: OOMEN, Raymond P.  
; APPLICANT: KLEIN, Michel H.  
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced Protease Act  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.125  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/482,816  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/296,149  
; FILING DATE: 26-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-494 MIS:vg  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 198 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-482-816-16

Query Match 58.1%; Score 39.5; DB 2; Length 198;  
Best Local Similarity 40.0%; Pred. No. 23;  
Matches 8; Conservative 3; Mismatches 0; Indels 9; Gaps 1;

```
Qy 1 CGVPA-----IQPVLS 11
||:|||:||||
Db 170 CGIPASQRSSLFRLQPILS 189

RESULT 7
US-08-296-149-16
; Sequence 16, Application US/08296149
; Patent No. 5939297
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COHEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,149
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-390
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 198 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-296-149-16

Query Match 58.1%; Score 39.5; DB 2; Length 198;
Best Local Similarity 40.0%; Pred. No. 23;
Matches 8; Conservative 3; Mismatches 0; Indels 9; Gaps 1;

Qy 1 CGVPA-----IQPVLS 11
||:|||:||||
Db 170 CGIPASQRSSLFRLQPILS 189

RESULT 8
US-08-801-499-16
; Sequence 16, Application US/08801499
; Patent No. 5962430
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COHEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced Protease Act
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
```

```
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,499
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/482,816
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-671 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 198 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-801-499-16

Query Match 58.1%; Score 39.5; DB 2; Length 198;
Best Local Similarity 40.0%; Pred. No. 23;
Matches 8; Conservative 3; Mismatches 0; Indels 9; Gaps 1;

Qy 1 CGVPA-----IQPVLS 11
||:|||:||||
Db 170 CGIPASQRSSLFRLQPILS 189

RESULT 9
US-08-615-271-16
; Sequence 16, Application US/08615271
; Patent No. 5981503
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COHEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
; TITLE OF INVENTION: PROTEASE ACTIVITY
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/615,271
```

```
; FILING DATE: 20-JUN-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-580
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1155
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 198 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-615-271-16

Query Match 58.1%; Score 39.5; DB 2; Length 198;
Best Local Similarity 40.0%; Pred. No. 23;
Matches 8; Conservative 3; Mismatches 0; Indels 9; Gaps 1;

QY 1 CGVPA-----IQPVLS 11
|||
Db 170 CGIPASQSRSLERQPLTS 189

RESULT 10
US-09-074-660-16
; Sequence 16, Application US/09074660
; Patent No. 6020183
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,660
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,167
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,660
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,167
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 198 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-615-271-16

Query Match 58.1%; Score 39.5; DB 2; Length 198;
Best Local Similarity 40.0%; Pred. No. 23;
Matches 8; Conservative 3; Mismatches 0; Indels 9; Gaps 1;
```

```
; LENGTH: 198 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-074-660-16

Query Match 58.1%; Score 39.5; DB 3; Length 198;
Best Local Similarity 40.0%; Pred. No. 23;
Matches 8; Conservative 3; Mismatches 0; Indels 9; Gaps 1;

QY 1 CGVPA-----IQPVLS 11
|||
Db 170 CGIPASQSRSLERQPLTS 189

RESULT 11
US-09-074-659-16
; Sequence 16, Application US/09074659
; Patent No. 6023342
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,659
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,167
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-730 MIS:jb
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 198 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-074-659-16

Query Match 58.1%; Score 39.5; DB 3; Length 198;
Best Local Similarity 40.0%; Pred. No. 23;
Matches 8; Conservative 3; Mismatches 0; Indels 9; Gaps 1;

QY 1 CGVPA-----IQPVLS 11
```

1



```
;
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/615,271
; FILING DATE: 20-JUN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-824
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 198 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-09-106-467-16
;
; Query Match 58.1%; Score 39.5; DB 4; Length 198;
; Best Local Similarity 40.0%; Pred No. 23;
; Matches 8; Conservative 3; Mismatches 0; Indels 9; Gaps 1;
;
; QY 1 CGVPA-----IQPVLS 11
;      ||||| :|||
; DB 170 CGIPASQSSLFRLQPILS 189
;
; RESULT 15
; US-08-681-129-2
; Sequence 2, Application US/08681129
; Patent No. 5738854
; GENERAL INFORMATION:
; APPLICANT: Mettenleiter, Thomas Cristoph
; TITLE OF INVENTION: Pseudorabies virus vaccine
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Organon Teknika Corporation
; STREET: 1330 Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/681,129
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/244,446
; FILING DATE: 02-JUN-1994
; CLASSIFICATION: 424
; APPLICATION NUMBER: EP 92.203.079.6
; FILING DATE: 06-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Gormley, Mary E.
; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 525 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Pseudorabies virus
;
;
; STRAIN: Kaplan
; US-08-681-129-2
;
; Query Match 57.4%; Score 39; DB 1; Length 525;
; Best Local Similarity 54.5%; Pred. No. 77;
; Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
;
; QY 2 GVPATQPVLSG 12
;      |:::|
; DB 187 GIPGVRPPLSG 197
;
; Search completed: February 12, 2003, 10:30:10
; Job time : 3.79403 secs
```

**THIS PAGE BLANK (USPTO)**



```
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 643
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (94)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (126)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (130)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (133)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (137)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (143)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-643

Query Match          92.6%; Score 63; DB 10; Length 146;
Best Local Similarity 92.3%; Pred. No. 0.001;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLSGL 13
Db 23 CGVPAIHPVLSGL 35

RESULT 3
US-09-888-615-96
; Sequence 96, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: FLOWMAN, GREGORY
; APPLICANT: WHYTE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARVDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 96
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-888-615-96

Query Match          92.6%; Score 63; DB 10; Length 263;
Best Local Similarity 92.3%; Pred. No. 0.0019;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLSGL 13
```

---

```
Db 19 CGVPAIHPVLSGL 31

RESULT 4
US-09-925-297-529
; Sequence 529, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: ROSEN ET AL.
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS AND ANTIBODIES
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 529
; LENGTH: 192
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-297-529

Query Match          72.1%; Score 49; DB 10; Length 192;
Best Local Similarity 72.7%; Pred. No. 0.31;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGVPAIQPVLS 11
Db 15 CGIPAIPKALS 25

RESULT 5
US-10-036-371-3
; Sequence 3, Application US/10036371
; Patent No. US20020141987A1
; GENERAL INFORMATION:
; APPLICANT: BJARNARSON, JON B.
; TITLE OF INVENTION: FISH SERINE PROTEINASES AND THEIR PHARMACEUTICAL AND
; FILE REFERENCE: 81691/284960
; CURRENT APPLICATION NUMBER: US/10/036,371
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: 09/411,688
; PRIOR FILING DATE: 1999-10-12
; PRIOR APPLICATION NUMBER: 5086/99
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Gadus sp.
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (3)
; OTHER INFORMATION: R or S
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (7)
; OTHER INFORMATION: S or Q
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (9)
; OTHER INFORMATION: V or Q
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (10)
; OTHER INFORMATION: I or V
US-10-036-371-3
```

```

Query Match      58.8%; Score 40; DB 12; Length 13;
Best Local Similarity 58.3%; Pred. No. 0.54;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CGVPAIQPVLSG 12
Db 1 CGXPAIXPXTG 12
      |||||
      :|

RESULT 6
US-09-764-869-944
; Sequence 944, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 944
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-869-944

Query Match      57.4%; Score 39; DB 10; Length 78;
Best Local Similarity 53.8%; Pred. No. 5.6;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CGVPAIQPVLSGL 13
Db 15 CGIPAGDPAASSL 27
      |||||
      :|

RESULT 7
US-09-862-540-34
; Sequence 34, Application US/09862540
; Publication No. US20030023992A1
; GENERAL INFORMATION:
; APPLICANT: Vogeli, Gabriel
; TITLE OF INVENTION: No. US20030023992A1el G Protein-Coupled Receptors
; FILE REFERENCE: 002520S1
; CURRENT APPLICATION NUMBER: US/09/862,540
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: 60/206,138
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/206,139
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/208,976
; PRIOR FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-862-540-34

Query Match      57.4%; Score 39; DB 9; Length 168;
Best Local Similarity 70.0%; Pred. No. 13;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 4 PAIQPVLSGL 13
Db 90 PELQPVLAGL 99
      |||||
      :|

RESULT 8
US-09-999-248-4
; Sequence 4, Application US/09999248
; Patent No. US20020176852A1
; GENERAL INFORMATION:
; APPLICANT: Lambeth, J. David
; APPLICANT: Cheng, Guangjie
; TITLE OF INVENTION: Mitogenic Oxygenase Regulators
; FILE REFERENCE: 05501-0180 43150-266489
; CURRENT APPLICATION NUMBER: US/09/999,248
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,305
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/251,364
; PRIOR FILING DATE: 2000-12-05
; PRIOR APPLICATION NUMBER: US 60/289,172
; PRIOR FILING DATE: 2001-05-07
; PRIOR APPLICATION NUMBER: US 60/289,537
; PRIOR FILING DATE: 2001-05-07
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-999-248-4

Query Match      55.9%; Score 38; DB 9; Length 565;
Best Local Similarity 58.3%; Pred. No. 70;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CGVPAIQPVLSG 12
Db 540 CGSPALAKVLKG 551
      |||||
      :|

RESULT 9
US-09-774-639-283
; Sequence 283, Application US/09774639
; Publication No. US20030003555A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 90 Human Secreted Proteins
; FILE REFERENCE: P2013P1
; CURRENT APPLICATION NUMBER: US/09/774,639
; CURRENT FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/244,112
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-02-04
; NUMBER OF SEQ ID NOS: 371
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 283
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-774-639-283

Query Match      54.4%; Score 37; DB 9; Length 24;
Best Local Similarity 63.6%; Pred. No. 3.4;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GVPAlQPVLSG 12
Db 10 GVPVHPAGSG 20
      |||||
      :|

RESULT 10
US-09-774-639-282
; Sequence 282, Application US/09774639
; Publication No. US20030003555A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 90 Human Secreted Proteins
; FILE REFERENCE: P2013P1
; CURRENT APPLICATION NUMBER: US/09/774,639
; CURRENT FILING DATE: 2001-07-09

```

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/244,112  
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-02-04  
; NUMBER OF SEQ ID NOS: 371  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 282  
; LENGTH: 48  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-774-639-282

Query Match 54.4%; Score 37; DB 9; Length 48;  
Best Local Similarity 63.6%; Pred. No. 7.2;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 GVPALQVLSG 12  
|||||  
Db 20 GVPAPHPAGS 30

RESULT 11  
US-09-864-761-36817  
; Sequence 36817, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecmica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 36817  
; LENGTH: 51

; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL049733.3  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.97  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.93  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.92  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.91  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.1  
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.98  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.95  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2  
; OTHER INFORMATION: EST\_HUMAN HIT: AW961297.1, EVALUATE 2.00e-03  
US-09-864-761-36817

Query Match 54.4%; Score 37; DB 10; Length 51;  
Best Local Similarity 50.0%; Pred. No. 7.6;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 GVPALQVLSGL 13  
|||||  
Db 10 GIPPIITPMPGL 21

RESULT 12  
US-09-864-761-46426  
; Sequence 46426, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecmica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30  
 ; PRIOR APPLICATION NUMBER: US 09/774,203  
 ; PRIOR FILING DATE: 2001-01-29  
 ; NUMBER OF SEQ ID NOS: 49117  
 ; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 46426  
 ; LENGTH: 52  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO AC010251.3  
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3  
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5  
 ; OTHER INFORMATION: EST\_HUMAN HIT: BE063509.1, EVALUATE 2.00e-04  
 US-09-864-761-46426

Query Match 54.4%; Score 37; DB 10; Length 52;  
 Best Local Similarity 63.6%; Pred. No. 7.8;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GVPAPQPVLS 11  
 Db 42 CTTPAIPVIS 52

RESULT 13  
 US-09-764-860-305  
 ; Sequence 305, Application US/09764860  
 ; Patent No. US20020094953A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PC008  
 ; CURRENT APPLICATION NUMBER: US/09/764,860  
 ; CURRENT FILING DATE: 2001-01-17  
 ; Prior application data removed - consult PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 1198  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 305  
 ; LENGTH: 99  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (40)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; NAME/KEY: SITE  
 ; LOCATION: (79)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; NAME/KEY: SITE  
 ; LOCATION: (81)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; NAME/KEY: SITE  
 ; LOCATION: (83)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 US-09-764-860-305

Query Match 54.4%; Score 37; DB 10; Length 99;  
 Best Local Similarity 58.3%; Pred. No. 16;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 GVPAPQPVLSGL 13  
 Db 74 GVPKLPXLXGL 85

RESULT 14  
 US-10-114-893-313  
 ; Sequence 313, Application US/10114893  
 ; Publication No. US20020193567A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Jacobs, Kenneth  
 ; APPLICANT: McCoy, John M.

; APPLICANT: LaVallie, Edward R.  
 ; APPLICANT: Collins-Racie, Lisa A.  
 ; APPLICANT: Evans, Cheryl  
 ; APPLICANT: Merberg, David  
 ; APPLICANT: Treacy, Maurice  
 ; APPLICANT: Bowman, Michael R.  
 ; APPLICANT: Spaulding, Vikki  
 ; APPLICANT: Carlin-Duckett, McKeough  
 ; APPLICANT: Kelleher, Kerry S.  
 ; APPLICANT: Genetics Institute, Inc.  
 ; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
 ; FILE REFERENCE: GI 6000-10A  
 ; CURRENT APPLICATION NUMBER: US/10/114,893  
 ; CURRENT FILING DATE: 2002-04-02  
 ; EARLIER APPLICATION NUMBER: 09/413,232  
 ; EARLIER FILING DATE: 1999-10-06  
 ; NUMBER OF SEQ ID NOS: 321  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 313  
 ; LENGTH: 125  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-114-893-313

Query Match 54.4%; Score 37; DB 9; Length 125;  
 Best Local Similarity 58.3%; Pred. No. 20;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GVPAPQPVLSGL 13  
 Db 52 GLPAPQPLTSL 63

RESULT 15  
 US-09-864-761-33801  
 ; Sequence 33801, Application US/09864761  
 ; Patent No. US20020048763A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharron G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; APPLICANT: Chen, Wensheng  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 ; FILE REFERENCE: Aeomica-X-1  
 ; CURRENT APPLICATION NUMBER: US/09/864,761  
 ; CURRENT FILING DATE: 2001-05-23  
 ; PRIOR APPLICATION NUMBER: US 60/180,312  
 ; PRIOR FILING DATE: 2000-02-04  
 ; PRIOR APPLICATION NUMBER: US 60/207,456  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: US 09/632,366  
 ; PRIOR FILING DATE: 2000-08-03  
 ; PRIOR APPLICATION NUMBER: GB 24263.6  
 ; PRIOR FILING DATE: 2000-10-04  
 ; PRIOR APPLICATION NUMBER: US 60/236,359  
 ; PRIOR FILING DATE: 2000-09-27  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662  
 ; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33801
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009237.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
; OTHER INFORMATION: EST\_HUMAN HIT: A0042813.1, EVALUE 2.00e-40
; OTHER INFORMATION: EST\_HUMAN HIT: BF345274.1, EVALUE 2.00e-45
; OTHER INFORMATION: SWISSPROT HIT: Q95157, EVALUE 9.00e-47
US-09-864-761-33801

Query Match 54.4% Score 37; DB 10; Length 147;
Best Local Similarity 70.0%; Pred.No: 24;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 PAIQPVLSGL 13
| : ||| |
Db 25 PELQPVLPGL 34

Search completed: February 12, 2003, 10:31:06
Job time : 2.9791 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:03:55 ; Search time 13.4328 Seconds  
(without alignments)  
198.395 Million cell updates/sec

Title: US-10-036-371-6

Perfect score: 112

Sequence: 1 IVNGEAVPGSWPQVSLQD 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002:\*

- 1: /SID32/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SID32/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SID32/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SID32/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SID32/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SID32/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SID32/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SID32/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SID32/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SID32/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SID32/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SID32/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SID32/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SID32/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SID32/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SID32/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SID32/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SID32/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SID32/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SID32/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SID32/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SID32/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SID32/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	112	100.0	20	AAB31578	Amino terminal seq
2	112	100.0	245	AA199596	Bovine chymotrypsin
3	109	97.3	20	AA193344	Bovine chymotrypsin
4	109	97.3	20	AA193195	N-terminal of a bo
5	109	97.3	20	AA190794	N-terminal of bovi
6	109	97.3	146	AA195419	Human pancreatic c
7	109	97.3	231	AA198504	Human chymotrypsin
8	109	97.3	263	AA198738	Amino acid sequenc
9	107	95.5	264	AA191711	Mouse serine prote
10	97	86.6	192	AA194077	Human pancreatic c

11	97	86.6	264	21	AAB11710	Human serine prote
12	93	83.0	31	17	AAW02588	PX, porcine pancre
13	93	83.0	242	7	AAP60061	Sequence of human
14	93	83.0	253	13	AAR29621	Porcine pancreatic
15	90	80.4	29	16	AAR84270	Serum calcium lowe
16	90	80.4	268	17	AAR90682	Rat caldesmon cont
17	88	78.6	242	7	AAP60060	Sequence of human
18	88	78.6	270	8	AAP70759	Human pancreatic t
19	88	78.6	270	23	AAU87691	Sequence of human
20	86	76.8	252	7	AAP60058	Human pancreas ela
21	86	76.8	269	7	AAP60062	Sequence of human
22	86	76.8	269	7	AAP61723	Human elastase II.
23	86	76.8	269	21	AAB54124	Human pancreatic c
24	86	76.8	273	21	AAB54243	Human pancreatic c
25	86	76.8	279	22	ABG20513	Novel human diagno
26	85	75.9	20	22	AAB31576	Amino terminal seq
27	84	75.0	283	21	AAE14615	Human prostate-ass
28	84	75.0	283	23	AAE14615	Human prostate-ass
29	84	75.0	393	22	AAU69962	Human prostate-ass
30	84	75.0	393	22	AAU69962	Human prostate-ass
31	84	75.0	393	23	AAU69962	Human prostate-ass
32	84	75.0	492	21	AB895422	Human TMPSR2 prot
33	84	75.0	492	21	AB895422	Human TMPSR2 prot
34	84	75.0	492	21	AAV92050	Human TMPSR2 prot
35	84	75.0	492	21	AAV92050	Human TMPSR2 prot
36	84	75.0	492	21	AAV92050	Human TMPSR2 prot
37	84	75.0	492	21	AAV92050	Human TMPSR2 prot
38	84	75.0	492	21	AAV92050	Human TMPSR2 prot
39	84	75.0	492	21	AAV92050	Human TMPSR2 prot
40	84	75.0	492	21	AAV92050	Human TMPSR2 prot
41	84	75.0	492	21	AAV92050	Human TMPSR2 prot
42	84	75.0	492	21	AAV92050	Human TMPSR2 prot
43	84	75.0	492	21	AAV92050	Human TMPSR2 prot
44	84	75.0	492	21	AAV92050	Human TMPSR2 prot
45	84	75.0	492	21	AAV92050	Human TMPSR2 prot

## ALIGNMENTS

RESULT 1  
AAB31578  
ID AAB31578 standard; peptide; 20 AA.  
XX AAB31578;  
AC AAB31578;  
XX 20-APR-2001 (first entry)  
DE Amino terminal sequence of bovine trypsin B chain.  
XX Fish; serine proteinase; pain; acute inflammation; chronic inflammation;  
KW arthritis; inflamed joint; bursitis; osteoarthritis; septic arthritis;  
KW rheumatoid arthritis; juvenile rheumatoid arthritis; fibromyalgia;  
KW systemic lupus erythematosus; phlebitis; tendinitis; rash; psoriasis;  
KW acne; eczema; facial seborrheic eczema; foreskin infection;  
KW athlete's foot; fistulae infection; ulcer; navel infection;  
KW scar; keloid; boil; wart; allergic itch; hemorrhoid; wound;  
KW fungal infection; autoimmune disease.  
XX Bos sp.  
OS Bos sp.  
XX WO200078332-A2.  
PN 28-DEC-2000.  
XX 15-JUN-2000; 2000WO-ISO0005.  
PF 18-JUN-1999; 99IS-0005086.  
PR (BJAR/) BJARNASON J B.  
PA Bjarnason JB;  
PI Bjarnason JB;

DR WPI; 2001-091493/10.

XX Fish serine proteinase, useful as a cosmetic, medicament for treating

PT eczema, psoriasis, arthritis, and in the manufacture of the medicament

PT for treating, preventing pathogenic diseases involving receptor

PT mediated binding -

XX Disclosure; Page 5; 38pp; English.

XX The specification describes a fish serine proteinase. The proteinases

CC are useful as medicaments, for treating and preventing a disease in a

CC human or an animal such as pain, acute inflammation, chronic

CC inflammation, arthritis, juvenile rheumatoid arthritis, osteoarthritis,

CC rheumatoid arthritis, systemic lupus erythematosus, phlebitis, tendinitis, rash,

CC fibromyalgia, eczema, facial seborrheic eczema, eczema of the hands,

CC psoriasis, acne, eczema, athlete's foot, fistulae infections,

CC face or neck, foreskin infections, navel infections in newborns, scars,

CC infected topical ulcers, wound infections in newborns, wounds, wound

CC keloids, boils, warts and allergic itch, hemorrhoids, immunological and

CC infections, wounds from burns, fungal infections and immunological and

CC autoimmune diseases. They are also useful for removing dead or peeling

CC skin from otherwise healthy skin, and for treating or preventing a

CC disease in which pathogenesis is caused by bacteria, virus, fungus,

CC parasite or a protozoan or a receptor mediated binding is involved.

CC The present sequence represents the amino terminal of bovine trypsin B

CC chain.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 112; DB 22; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.7e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IVNGEEAVPGSWPQVSLQD 20

Db 1 IVNGEEAVPGSWPQVSLQD 20

RESULT 2

AAV99596

ID AAV99596 standard; protein; 245 AA.

XX AAV99596;

XX 13-SEP-2000 (first entry)

XX Bos taurus.

XX Bovine chymotrypsinogen A.

XX Bovine; plasminogen activator; cardiant; thrombolytic;

XX heart attack; stroke; blood clotting disorder.

XX OS

XX WO200032759-A1.

XX 08-JUN-2000.

XX 06-MAY-1999; 99WO-US09991.

XX 02-DEC-1998; 98US-0110588.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX Lin X, Zhang XC, Tang JJJ;

XX WPI; 2000-422975/36.

XX Polypeptide with plasminogen activator activity useful as thrombolytic

PT agent for treating blood clot disorders e.g. heart attack, comprises 10

PT amino acid peptide fragment for recognition or activation of

PT plasminogen -

XX Disclosure; Page 35-36; 41pp; English.

XX The present sequence is bovine chymotrypsinogen. It was included in

CC a review of sequence homologies of several plasminogen activators.

CC Plasminogen is the principal serine protease zymogen in the

CC extracellular fluids of vertebrates. Its active form, plasmin, is

CC implicated in pericellular proteolysis associated with a wide range of

CC physiological and pathological processes. Plasminogen expression is

CC regulated by plasminogen activators which hydrolyse a peptide bond in

CC plasminogen to convert it to plasmin or form tight binding complexes

CC with plasminogen to spontaneously convert it to plasmin. The sequence

CC homology analysis has identified a six amino acid peptide involved in

CC plasminogen activation. This peptide is particularly useful when inserted

CC between amino acid residues 644 and 645 of full length human

CC plasminogen. Novel plasminogen activators have been made based upon the

CC plasminogen activation/recognition site of plasminogen binding proteins.

CC The polypeptides are useful in preparing thrombolytic agents for treating

XX blood clotting disorders such as heart attack.

XX SQ Sequence 245 AA;

Query Match 100.0%; Score 112; DB 21; Length 245;

Best Local Similarity 100.0%; Pred. No. 2.7e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IVNGEEAVPGSWPQVSLQD 20

Db 16 IVNGEEAVPGSWPQVSLQD 35

RESULT 3

AAV33344

ID AAV33344 standard; peptide; 20 AA.

XX AAV33344;

XX 29-NOV-1999 (first entry)

XX Bovine chymotrypsin N-terminal fragment.

XX Multifunctional enzyme; treatment; acne; eczema; chymotrypsin; trypsin;

XX collagenase; elastase; exopeptidase; cell surface receptor; anti-acne;

XX antiseborrheic; dermatological; anti-eczema; adhesion molecule; ICAM-1;

XX CD54; ICAM-2; VCAM-1; CD4; CD8; CD28; CD29D; CD31; CD44; CD49; CD62L;

XX CD102; GMI ceramide; krill; bovine.

XX OS

XX Bos taurus.

XX US9598406-A.

XX 28-SEP-1999.

XX 08-FEB-1996; 96US-0600273.

XX 22-NOV-1994; 94US-0388501.

XX 08-FEB-1995; 95US-0385540.

XX 07-JUN-1995; 95US-0486820.

XX (PHAI-) PHAIRSON MEDICAL INC.

XX Lindblom R, Kay J, Franklin RL, De Paire JR;

XX WPI; 1999-561004/47.

XX Treating acne and eczema using a krill-derived multifunctional enzyme -

XX Disclosure; Column 21-22; 42pp; English.

XX This invention describes a novel method for treating acne and eczema

CC using a krill-derived multifunctional enzyme (I) which comprises 2 or

CC more of the activities of chymotrypsin, trypsin, collagenase, elastase

CC or exopeptidase and is reactive with cell surface receptors such as

CC proteins or glycoproteins. The product of the invention have

CC antiseborrheic, anti-acne, dermatological and anti-eczema activity.

CC (1) removes or inactivates cell surface receptors (proteins and  
 CC glycoproteins) and adhesion molecules such as ICAM-1 (i.e. CD54), ICAM-2,  
 CC VCAM-1, CD4, CD8, CD28, CD29D, CD31, CD49, CD62L CD102 and the  
 CC asialo GM1 ceramide. This sequence represents the N-terminal fragment of  
 CC bovine chymotrypsin which is used to describe the method of the  
 CC invention.

XX SQ Sequence 20 AA;

Query Match 97.3%; Score 109; DB 20; Length 20;  
 Best Local Similarity 95.0%; Pred. No. 4.6e-09;  
 Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPQVSLQD 20  
 Db 1 IVNGEDAVPGSWPQVSLQD 20  
 |||||:|||||:|||||:|||||

#### RESULT 4

AAAY93935  
 ID AAY93935 standard; peptide; 20 AA.

AC AAY93935;

DT 03-OCT-2000 (first entry)

DE N-terminal of a bovine chymotrypsin hydrolase enzyme.

KW Transplantation rejection; hydrolase; graft versus host disease;  
 KW cell surface adhesion molecule; immune reaction; inflammation; shock;  
 KW tumour metastasis; autoimmune disease;  
 KW Krill derived multifunctional enzyme.

OS Bos sp.

PN WO200038708-A1.

XX PD 06-JUL-2000.

XX PF 23-DEC-1999; 99WO-US30818.

XX PR 24-DEC-1998; 98US-0114147.

XX PA (PHAI-) PHAIRSON MEDICAL INC.

XX PI Franklin RL, St Pierre Y;

XX DR WPI; 2000-452301/39.

XX PT Preventing or ameliorating transplantation rejection reactions using  
 PT hydrolase enzymes

XX PS Disclosure; Page 26; 66pp; English.

CC The specification describes a method for preventing or ameliorating  
 CC transplantation rejection reactions for transplantation of immune cells  
 CC or other tissues. The method comprises treating a source of immune cells  
 CC with a hydrolase or hydrolase mixture and administering the treated  
 CC immune cells to a recipient animal. The hydrolase especially has a  
 CC preference for removing, destroying, inactivating or disabling at least  
 CC one of CD4, CD8, CD25, CD28, ICAM-1 (CD54), CD152, an integrin, CD154,  
 CC CD40 and CD80 in contrast to removing, destroying, inactivating or  
 CC disabling Tcr. The methods are useful for preventing graft versus host  
 CC disease by using hydrolase enzymes to remove the cell surface adhesion  
 CC molecules which are involved in triggering the immune reactions involved  
 CC in the diseases. The methods are used for treating or preventing  
 CC cell-cell or cell-virus adhesion syndrome comprising inflammation, shock,  
 CC tumour metastases, autoimmune disease, transplantation rejection  
 CC reactions or microbial infections. The present sequence represents the  
 CC N-terminal of a hydrolase, which may be used in the method of the  
 CC invention.

XX SQ Sequence 20 AA;

Query Match 97.3%; Score 109; DB 21; Length 20;  
 Best Local Similarity 95.0%; Pred. No. 4.6e-09;  
 Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPQVSLQD 20  
 Db 1 IVNGEDAVPGSWPQVSLQD 20  
 |||||:|||||:|||||:|||||

#### RESULT 5

AAE07941  
 ID AAE07941 standard; peptide; 20 AA.

AC AAE07941;

DT 01-NOV-2001 (first entry)

DE N-terminal of bovine serine multifunctional hydrolase.

KW Multifunctional hydrolase; rejection reaction; extra-corporeal;  
 KW therapy; graft versus host disease; transplantation rejection;  
 KW autoimmune disease; microbial infection; immune disorder; cytostatic;  
 KW cystic fibrosis; chronic obstructive pulmonary disease; COPD;  
 KW atherosclerosis; cancer; asthma; septic shock; toxic shock syndrome;  
 KW conjunctivitis; reperfusion injury; pain; immunosuppressive;  
 KW antibacterial; vasotropic; bovine.

OS Bos sp.

PN US6232088-B1.

XX PD 15-MAY-2001.

XX PF 24-DEC-1998; 98US-0220731.

XX PR 08-FEB-1995; 95US-0385540.

XX PR 07-JUN-1995; 95US-0486820.

XX PR 08-FEB-1996; 96US-0600273.

XX PA (PHAI-) PHAIRSON MEDICAL INC.

XX PI Franklin RL, St Pierre Y;

XX DR WPI; 2001-450051/48.

XX PT Ameliorating, preventing or treating immune rejection reactions, such  
 PT as graft versus host disease, autoimmune disease, asthma, cancer, by  
 PT extra-corporeally treating donor tissue with hydrolase such as a  
 PT protease

XX PS Disclosure; Column 15-16; 27pp; English.

CC The present invention relates to a method for ameliorating  
 CC transplantation rejection reaction. The method comprises treating  
 CC extra-corporeally donor tissue or donor source of immune cells with  
 CC a rejection ameliorating effective amount of a hydrolase enzyme. The  
 CC method is useful for ameliorating, treating or preventing immune  
 CC rejection reactions such as graft versus host diseases, organ or tissue  
 CC transplantation rejection, autoimmune disease and associated conditions,  
 CC microbial infection, immune disorder, cystic fibrosis, chronic  
 CC obstructive pulmonary disease (COPD), atherosclerosis, cancer, asthma,  
 CC septic shock, toxic shock syndrome, conjunctivitis, reperfusion injury  
 CC and pain in humans. The present sequence is the N-terminal peptide  
 CC of bovine serine multifunctional hydrolase, used in the  
 CC exemplification of the invention.

XX SQ Sequence 20 AA;

Query Match 97.3%; Score 109; DB 22; Length 20;  
 Best Local Similarity 95.0%; Pred. No. 4.6e-09;  
 Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 IVNGEEAVPGSWPQVSLQD 20  
 |||||:|||||  
 Db 38 IVNGEDAVPGSWPQVSLQD 57

RESULT 7  
 AAB98504  
 ID AAB98504 standard; Protein; 231 AA.  
 XX  
 AC AAB98504;  
 XX  
 DT 03-AUG-2001 (first entry)  
 XX

DE Human chymotrypsin serine protease catalytic domain.  
 KW Human; TADG-15; cytostatic; vaccine; ovarian tumour; cancer;  
 KW tumour antigen-derived gene 15; serine protease; chymotrypsin.  
 XX Homo sapiens.  
 OS  
 XX WO200129056-A1.  
 XX  
 XX 26-APR-2001.  
 XX  
 XX 20-OCT-2000; 2000WO-US29095.  
 XX  
 XX 20-OCT-1999; 99US-0421213.  
 PR  
 XX (UYAR-) UNIV ARKANSAS.  
 PA  
 XX O'Brien TJ, Tanimoto H;  
 PI  
 XX WPI; 2001-381031/40.  
 DR  
 XX Novel extracellular serine protease, termed tumor antigen-derived gene  
 PT 15 protein overexpressed in carcinomas and DNA encoding it, for  
 PT diagnosis, treatment, prevention of cancer, particularly breast,  
 PT ovarian cancer -  
 XX

PS Example 10; Fig 1; 130pp; English.  
 XX  
 CC The present invention relates to human tumour antigen-derived gene 15  
 CC (TADG-15) protein and coding sequence (see AAH3601 and AAB98500).  
 CC (TADG-15) is an extracellular serine protease. It was found that TADG-15 is  
 CC over-expressed in ovarian tumours. TADG-15 protein or its fragments of  
 CC 9-20 residues that lack TADG-15 protease activity are useful for  
 CC vaccinating an individual against TADG-15, having, suspected of having or  
 CC at risk of getting cancer. Furthermore, the TADG-15 gene can be used as a  
 CC diagnostic or therapeutic target in cancer. The present sequence was used  
 CC in a sequence homology alignment with the catalytic domain of TADG-15.  
 XX

SQ Sequence 231 AA;  
 Query Match 97.3%; Score 109; DB 22; Length 231;  
 Best Local Similarity 95.0%; Pred. No. 6.8e-08;  
 Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 IVNGEEAVPGSWPQVSLQD 20  
 |||||:|||||  
 Db 2 IVNGEDAVPGSWPQVSLQD 21

RESULT 8  
 AAU82738  
 ID AAU82738 standard; Protein; 263 AA.  
 XX  
 AC AAU82738;  
 XX  
 DT 23-APR-2002 (first entry)  
 XX

DE Amino acid sequence of novel human protease #37.  
 XX

Oy 1 IVNGEEAVPGSWPQVSLQD 20  
 |||||:|||||  
 Db 1 IVNGEDAVPGSWPQVSLQD 20

RESULT 6  
 AAB54191  
 ID AAB54191 standard; Protein; 146 AA.  
 XX  
 AC AAB54191;  
 XX  
 DT 09-MAR-2001 (first entry)  
 XX

DE Human pancreatic cancer antigen protein sequence SEQ ID NO:643.  
 KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;  
 KW detection; diagnosis; identification; cytostatic; neuroprotective;  
 KW nontropic; immunomodulatory; relaxant; contraceptive; gynaecological;  
 KW antiinflammatory; cardiant; gene therapy; chromosome mapping;  
 KW linkage analysis; tissue identification; tissue typing; forensic;  
 KW neural; immune system; muscular; reproductive; gastrointestinal;  
 KW pulmonary; cardiovascular; renal; proliferative.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200055320-A1.  
 XX  
 XX 21-SEP-2000.  
 PD  
 XX 08-MAR-2000; 2000WO-US05989.  
 XX  
 XX 12-MAR-1999; 99US-0124270.  
 PR  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Rosen CA, Ruben SM;  
 PI  
 XX WPI; 2000-579444/54.  
 DR  
 XX N-PSDB; AAC98956.

XX New nucleic acid that is a pancreatic cancer antigen for preventing,  
 PT treating, or ameliorating a medical condition, particular pancreatic  
 PT cancer, or for use in assays for diagnosing a pathological condition -  
 XX  
 PS Claim 11; Page 1081; 1379pp; English.  
 XX  
 CC AAC98773 to AAC99231 encode the human pancreatic cancer associated  
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to  
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,  
 CC neuroprotective, nontropic, immunomodulatory, relaxant, contraceptive,  
 CC gynaecological, cardiant and antiinflammatory activities, and can be used  
 CC in gene therapy. The polynucleotide and proteins can be used for  
 CC preventing, treating, or ameliorating a medical condition or in assays  
 CC for diagnosing a pathological condition or a susceptibility to one in a  
 CC subject. Binding partners to the proteins and the activity of the  
 CC proteins can be identified. The pancreatic cancer antigens can be used to  
 CC detect, treat or prevent pancreatic disorders, especially cancer.  
 CC Agonists and antagonists to the antigens can be screened for. The  
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
 CC acid hybridisation probes that can be used in chromosome mapping, linkage  
 CC analysis, tissue identification and/or typing and a variety of forensic  
 CC and diagnostic methods. The proteins can be used to generate antibodies  
 CC which are used to purify, detect and target the polypeptides, including  
 CC both in vivo and in vitro diagnostic and therapeutic methods. The  
 CC proteins can be used to treat or prevent neural, immune system, muscular,  
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
 CC sequences used in the exemplification of the present invention.  
 XX

SQ Sequence 146 AA;  
 Query Match 97.3%; Score 109; DB 21; Length 146;  
 Best Local Similarity 95.0%; Pred. No. 4.1e-08;



KW detection; diagnosis; identification; cytostatic; neuroprotective;  
KW neotropic; immunomodulatory; relaxant; contraceptive; gynaecological;  
KW antifungal; cardiant; gene therapy; chromosome mapping;  
KW linkage analysis; tissue identification; tissue typing; forensic;  
KW neural; immune system; muscular; reproductive; gastrointestinal;  
KW pulmonary; cardiovascular; renal; proliferative.  
XX  
OS Homo sapiens.  
XX  
PN WO200055320-A1.  
XX  
PD 21-SEP-2000.  
XX  
PP 08-MAR-2000; 2000WO-US05989.  
XX  
PR 12-MAR-1999; 99US-0124270.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Ruben SM;  
XX  
DR WPI; 2000-579444/54.  
DR N-PSDB; AAC98842.  
XX  
XX New nucleic acid that is a pancreatic cancer antigen for preventing,  
PT treating, or ameliorating a medical condition, particular pancreatic  
PT cancer, or for use in assays for diagnosing a pathological condition -  
XX  
PS Claim 11; Page 966; 1379pp; English.  
XX  
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated  
CC proteins, called pancreatic cancer antigens, given in AAB54008 to  
CC AAB54466. The human pancreatic cancer antigens have cytosolic,  
CC neuroprotective, neotropic, immunomodulatory, relaxant, contraceptive,  
CC gynaecological, cardiant and antiinflammatory activities, and can be used  
CC in gene therapy. The polynucleotide and proteins can be used for  
CC preventing, treating, or ameliorating a medical condition or in assays  
CC for diagnosing a pathological condition or a susceptibility to one in a  
CC subject. Binding partners to the proteins and the activity of the  
CC proteins can be identified. The pancreatic cancer antigens can be used to  
CC detect, treat or prevent pancreatic disorders, especially cancer.  
CC Agonists and antagonists to the antigens can be screened for. The  
CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
CC acid hybridisation probes that can be used in chromosome mapping, linkage  
CC analysis, tissue identification and/or typing and a variety of forensic  
CC and diagnostic methods. The proteins can be used to generate antibodies  
CC which are used to purify, detect and target the polypeptides, including  
CC both in vivo and in vitro diagnostic and therapeutic methods. The  
CC proteins can be used to treat or prevent neural, immune system, muscular,  
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
CC sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 192 AA;  
Query Match 86.6%; Score 97; DB 21; Length 192;  
Best Local Similarity 90.0%; Pred. No. 3e-06; 2; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Oy 1 IVNGEEAVPGSWPQVSLQD 20  
Db 30 IVNGENAVLGSWPQVSLQD 49  
RESULT 11  
AAB11710  
ID AAB11710 standard; Protein; 264 AA.  
XX  
AC AAB11710;  
XX  
XX 23-OCT-2000 (first entry)  
XX  
XX Human serine protease BSSP5 (hBSSP5) SEQ ID NO:2.  
DE

XX BSSP5, serine protease; human; hBSSP5; mouse; mBSSP5; brain;  
KW diagnostic marker; antibody; transgenic animal; Alzheimer's disease;  
KW epilepsy; cancer; inflammation; infertility; pancreatitis;  
KW prostatic hypertrophy.  
XX  
OS Homo sapiens.  
XX  
PN WO200031243-A1.  
XX  
PD 02-JUN-2000.  
XX  
PP 19-NOV-1999; 99WO-JP06473.  
XX  
PR 20-NOV-1998; 98JP-0347806.  
XX  
PA (FUSO) FUSO PHARM IND LTD.  
XX  
PI Uemura H, Okui A, Kominami K, Yamaguchi N, Mitsui S;  
XX  
DR WPI; 2000-400058/34.  
DR N-PSDB; AAA61733.  
XX  
XX Serine proteases BSSP5, useful in detecting homologs, mutants and  
PT polymorphic variants as markers for diagnosis of e.g. Alzheimer's  
PT disease, epilepsy, cancer and inflammation, using blood, urine,  
PT pancreas or other tissues -  
XX  
PS Claim 1; Page 51-52; 70pp; Japanese.  
XX  
XX The invention relates to novel serine proteases designated BSSP5  
CC (AAB11710-B11711), and to nucleic acids encoding them (AAA61733-A61734).  
CC The invention also relates to vectors and transformants comprising BSSP5  
CC nucleic acids; transgenic animals in which the expression level of BSSP5  
CC can be varied; and an mBSSP5 knockout mouse. The invention additionally  
CC encompasses anti-BSSP5 antibodies and methods of production of such  
CC antibodies, methods of BSSP5 detection using the antibodies, and the  
CC use of BSSP5 proteins or fragments as diagnostic markers for certain  
CC medical conditions, e.g., pancreatitis. A method for detecting  
CC pancreatitis comprising measuring BSSP5 concentration in the blood or  
CC urine, and a pancreatitis diagnostic agent containing an anti-BSSP5  
CC antibody is also disclosed. Nucleotides encoding BSSP5 were initially  
CC isolated in a human brain cDNA library using degenerate PCR primers  
CC (AAA61744-A61745) based on conserved regions of serine proteases. The  
CC BSSP5 serine proteases and nucleotides encoding them are useful in  
CC detecting homologues, mutants and polymorphic variants in biological  
CC samples (e.g., blood, urine, brain, prostate gland, placenta, testis,  
CC pancreas and spleen) as diagnostic markers for conditions such as  
CC Alzheimer's disease, epilepsy, cancer, inflammation, infertility,  
CC pancreatitis and prostatic hypertrophy. Sequence AAB11710 represents  
CC human BSSP5 (hBSSP5), and sequence AAB11711 represents murine BSSP5  
CC (mBSSP5).  
XX  
SQ Sequence 264 AA;  
Query Match 86.6%; Score 97; DB 21; Length 264;  
Best Local Similarity 90.0%; Pred. No. 4.2e-06; 0; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Oy 1 IVNGEEAVPGSWPQVSLQD 20  
Db 34 IVNGENAVLGSWPQVSLQD 53  
RESULT 12  
AAW02588  
ID AAW02588 standard; peptide; 31 AA.  
XX  
AC AAW02588;  
XX  
XX 12-FEB-1997 (first entry)  
XX  
XX PX, porcine pancreas-derived factor N-terminal peptide.  
DE

XX Pig; bone; calcium; resorption inhibition; formation; osteoporosis;  
 KW Paget's disease; fracture repair; bone defect; osteopetrosis;  
 KW metastasis; osteoblast; osteoclast.  
 XX

OS Sus scrofa.  
 PN WO9619501-A1.  
 XX

XX 27-JUN-1996.  
 PD

XX 19-DEC-1995; 95WO-US16826.  
 PF

XX 20-DEC-1994; 94US-0363092.  
 PR

XX (OSTE-) OSTEOSA INC.  
 PS

XX Izbicka E, Mundy GR, Yoneda T;  
 PI

XX WPI; 1996-309523/31.  
 DR

XX New isolated pancreatic-derived factor - which inhibits bone  
 PT resorption and increases bone formation, used to develop prods. for  
 PT diagnosis and therapy  
 PT

XX Claim 1; Fig 14; 53pp; English.  
 PS

XX AA02588 is the N-terminal portion of a factor derived from pig  
 CC pancreatic tissue. The factor was designated PX. PX inhibits the  
 CC action of osteoclasts and stimulates the action of osteoblasts,  
 CC hence PX inhibits bone resorption and stimulates bone formation. The  
 CC new factor also stimulates the proliferation of human MG-63  
 CC osteosarcoma cells (osteoblast-like cells). The factor is useful for  
 CC treating bone loss, e.g. due to osteoporosis or Paget's disease. The  
 CC factor may also be used to hasten bone fracture repair. Antibodies  
 CC and antagonists of PX can be used to treat osteopetrosis and  
 CC osteoblastic metastasis  
 CC

XX Sequence 31 AA;  
 SQ

Query Match 83.0%; Score 93; DB 17; Length 31;  
 Best Local Similarity 84.2%; Pred. No. 1.5e-06;  
 Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPQVSLQ 19  
 :||||:|||||  
 Db 1 VVNGEDAVPYSPWQVSLQ 19

RESULT 13  
 AAP60061  
 ID AAP60061 standard; Protein; 242 AA.  
 XX  
 AC AAP60061;  
 XX

DT 06-SEP-1991 (first entry)  
 DE

XX Sequence of human pancreatic elastase IIIB.  
 XX

XX Enzyme; serum lipoprotein metabolism.  
 KW

OS Homo sapiens.  
 XX

PN EP198645-A.  
 XX

XX 22-OCT-1986.  
 PD

XX 07-APR-1986; 86EP-0302557.  
 PF

XX 23-OCT-1985; 85JP-0236686.  
 PR

XX 05-APR-1985; 85JP-0072308.  
 PR

XX 27-APR-1985; 85JP-0091986.  
 PR

XX 26-JUL-1985; 85JP-0163964.  
 PR

PR 02-DEC-1985; 85JP-0271128.  
 XX  
 PA (SANY ) SANKYO KK.  
 XX

PI Takiguchi Y, Tani T, Kawashima I, Eurukawa H, Ohmine T;  
 PI Ohsumi J;  
 XX

XX WPI; 1986-280300/43.  
 DR

XX N-PSDB; AAN60075.  
 DR

XX New genetically engineered human pancreatic elastase - obid.  
 PT using hosts modified DNA coding for enzyme  
 PT

XX Claim 2; Page 8; 45pp; English.  
 PS

XX The genetically engineered prod. can eliminate the dependency on  
 CC human pancreas supplies for the elastase, and avoids antibody  
 CC formation and possibility of anaphylaxis using porcine elastase.  
 CC

XX Sequence 242 AA;  
 SQ

Query Match 83.0%; Score 93; DB 7; Length 242;  
 Best Local Similarity 84.2%; Pred. No. 1.5e-05;  
 Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPQVSLQ 19  
 :||||:|||||  
 Db 1 VVNGEDAVPYSPWQVSLQ 19

RESULT 14  
 AAR29621  
 ID AAR29621 standard; Protein; 253 AA.  
 XX  
 AC AAR29621;  
 XX

DT 30-APR-1993 (first entry)  
 DE

XX Porcine pancreatic elastase III.  
 XX

XX Pig; swine; bile acid; gall bladder; bile acid secretion promoters;  
 KW liver function improvers; 88.  
 KW

XX Sus scrofa domestica.  
 OS

XX JP04325090-A.  
 PN

XX 13-NOV-1992.  
 PD

XX 23-APR-1991; 91JP-0092069.  
 PF

XX 23-APR-1991; 91JP-0092069.  
 PR

XX (SANY ) SANKYO CO LTD.  
 PA

XX WPI; 1992-428826/52.  
 DR

XX N-PSDB; AAQ31724.  
 DR

XX Pig pancreas elastase protein - used in bile acid secretion  
 PT promoters and liver function improvers  
 PT

XX Claim 1; Page 8; 10pp; Japanese.  
 PS

XX This sequence represents porcine pancreatic elastase III. It was  
 CC expressed in E. Coli YA21 using the expression vector pSE001.  
 CC It may be used in bile acid secretion promoters and liver function  
 CC improvers.  
 CC

XX Sequence 253 AA;  
 SQ

Query Match 83.0%; Score 93; DB 13; Length 253;  
 Best Local Similarity 84.2%; Pred. No. 1.5e-05;  
 Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Wed Feb 12 11:59:41 2003

QY 1 IVNGEAVPGSNPWQVSLQ 19  
Db 12 VVNGEDAVFYSNPNWQVSLQ 30

RESULT 15  
AAR84270  
ID AAR84270 standard; peptide; 29 AA.  
XX  
AC AAR84270;  
XX  
DT 11-APR-1996 (first entry)  
XX  
DE Serum calcium lowering factor N-terminal sequence.  
XX  
KW Serum calcium lowering factor; rat; pancreas; calcium; mouse;  
KW bone disease; osteoporosis; primary hyperthyroidism;  
KW continuous hyperthyroidism; hypercalcaemia; malignant tumour.  
XX  
OS Rattus rattus.  
XX  
PN JP07215997-A.  
XX  
PD 15-AUG-1995.  
XX  
PF 01-FEB-1994; 94JP-0027578.  
XX  
PR 01-FEB-1994; 94JP-0027578.  
XX  
PA (CHUS ) CHUGAI PHARM CO LTD.  
XX  
XX WPI; 1995-317483/41.  
XX  
XX Serum calcium lowering factor obtd. from rat pancreas - useful as  
PT drug for treatment of various bone disease e.g. osteoporosis,  
PT hyperthyroidism and hypercalcaemia.  
XX  
PS Claim 2; Page 2; 6pp; Japanese.  
XX  
XX This sequence represents the N-terminal sequence of a serum calcium  
CC lowering factor derived from rat pancreas. The factor has a  
CC molecular weight of 30000 measured by SDS-PAGE and causes a decrease  
CC in serum levels of calcium in mice in a dose dependent manner. The  
CC factor may be used for the prevention and treatment of various bone  
CC diseases such as osteoporosis, primary and continuous hyperthyroidism  
CC and hypercalcaemia accompanying malignant tumours.  
XX  
SQ Sequence 29 AA;

Query Match 80.4%; Score 90; DB 16; Length 29;  
Best Local Similarity 78.9%; Pred. No. 3.8e-06;  
Matches 15; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSNPWQVSLQ 19  
Db 1 VVNGEDAVFYSNPNWQVSLQ 30

Search completed: February 12, 2003, 10:22:26  
Job time : 14.4328 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:17:50 ; Search time 5.01493 Seconds  
(without alignments)  
383.393 Million cell updates/sec

Title: US-10-036-371-6

Perfect score: 112

Sequence: 1 IVNGEAVPGSWPQVSLQD 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73.\*

1: Pir1.\*  
2: Pir2.\*  
3: Pir3.\*  
4: Pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112	100.0	245	1 KYBOA	chymotrypsin (EC 3
2	109	97.3	245	1 KYBOB	chymotrypsin (EC 3
3	109	97.3	263	2 A21195	chymotrypsin (EC 3
4	109	97.3	263	2 A31299	chymotrypsin (EC 3
5	108	96.4	263	1 KYRTB	chymotrypsin (EC 3
6	102	91.1	20	2 B61333	chymotrypsin (EC 3
7	98	87.5	126	2 A23473	chymotrypsin-like
8	98	87.5	244	2 S72219	chymotrypsin B - A
9	97	86.6	264	2 I38136	chymotrypsin-like
10	96	85.7	263	2 B3257	chymotrypsin (EC 3
11	93	83.0	31	2 B3257	pancreatic elastase
12	93	83.0	270	2 B29934	pancreatic elastase
13	88	78.6	270	2 A29934	pancreatic elastase
14	86	76.8	23	2 P00036	serine proteinase
15	86	76.8	269	2 B26823	pancreatic elastase
16	84	75.0	271	1 E3RT2	pancreatic elastase
17	83	74.1	269	2 A26823	pancreatic elastase
18	83	74.1	271	2 A26823	pancreatic elastase
19	82	73.2	268	2 JQ473	pancreatic elastase
20	81	72.3	269	2 C26823	pancreatic elastase
21	80	71.4	268	2 S68826	pancreatic elastase
22	80	71.4	268	2 S68825	pancreatic elastase
23	77	68.8	1524	2 T30337	polyprotein - Afri
24	76	67.9	421	2 S29599	acrosin (EC 3.4.21
25	75	67.0	37	2 S02176	procarboxypeptidase
26	73	65.2	240	1 C9R0A3	acrosin (EC 3.4.21
27	73	65.2	415	1 A34170	chymotrypsin (EC 3
28	72	64.3	28	2 A16529	chymotrypsin (EC 3
29	70	62.5	1004	2 T30338	oviductin (EC 3.4.

## ALIGNMENTS

## RESULT 1

## KYBOA

chymotrypsin (EC 3.4.21.1) A precursor - bovine

N:Alternate names: chymotrypsinogen A

C:Species: Bos primigenius taurus (cattle)

C:Date: 07-May-1981 #sequence\_revision 07-May-1981 #text\_change 07-May-1999

C:Accession: A90235; A93158; S29650; A00952

R:Brown, J.R.; Hartley, B.S.

Biochem. J. 101, 214-228, 1966

A:Title: Location of disulphide bridges by diagonal paper electrophoresis. The disulphide

A:Reference number: A90235; MUID:67181721; PMID:5971763

A:Accession: A90235

A:Molecule type: protein

A:Residues: 1-101, N', 103-245 <BRO>

R:Blow, D.M.; Birktoft, J.J.; Hartley, B.S.

Nature 221, 337-340, 1969

A:Title: Role of a buried acid group in the mechanism of action of chymotrypsin.

A:Reference number: A93158; MUID:69106266; PMID:5764436

A:Contents: annotation; revision to residue 102

R:Meloun, B.; Klueh, I.; Kostka, V.; Moravsek, L.; Prusik, Z.; Vanacek, J.; Keil, B.; Sorm

Biochim. Biophys. Acta 130, 543-546, 1966

A:Title: Covalent structure of bovine chymotrypsinogen A.

A:Reference number: A90572; MUID:67183948; PMID:5972866

A:Accession: A93158

A:Molecule type: protein

A:Residues: 1-101, N', 103-245 <MEL>

A:Note: disulfide bonds were determined

R:Cutruzzola, F.; Ascenzi, P.; Barra, D.; Bolognesi, M.; Menegatti, E.; Sarti, P.; Schnei

Biochim. Biophys. Acta 1161, 201-208, 1993

A:Title: Selective oxidation of Met-192 in bovine alpha-chymotrypsin. Effect on catalytic

A:Reference number: S29650; MUID:93160238; PMID:8431470

A:Accession: S29650

A:Molecule type: protein

A:Residues: 1-127, 16-27, 149-160, 181-200 <CUT>

R:Smillie, L.B.; Hartley, B.S.

Biochem. J. 101, 232-241, 1966

A:Title: Histidine sequences in the active centres of some 'serine' proteinases.

A:Reference number: A90236; MUID:67181723; PMID:5971785

A:Contents: annotation; active site

R:Birktoft, J.J.; Blow, D.M.; Henderson, R.; Steitz, T.A.

Philos. Trans. R. Soc. Lond. B257, 67-76, 1970

A:Title: The structure of alpha-chymotrypsin.

A:Reference number: A93754

A:Contents: annotation; X-ray crystallography

C:Comment: Chymotrypsinogens are synthesized in the acinar cells of pancreas.

delta-chymotrypsin; further chymotryptic cleavage liberates the dipeptide Thr-147 and Asn

d Asn-148 directly from chymotrypsinogen, which leads to the degraded form neochymotryps

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen

F:1-245/Product: chymotrypsinogen #status experimental <ZYM>

F:1-13,16-146,149-245/Product: alpha-chymotrypsin #status experimental <MPT>

collagenolytic pro  
acrosin (EC 3.4.21  
acrosin (EC 3.4.21  
acrosin (EC 3.4.21  
plasma kallikrein  
pancreatic elastase  
trypsin (EC 3.4.2  
mast cell tryptase  
trypsin (EC 3.4.2  
trypsin (EC 3.4.2  
trypsin (EC 3.4.2  
trypsin (EC 3.4.2  
trypsin (EC 3.4.2  
mast cell proteinase  
trypsin (EC 3.4.2  
trypsin (EC 3.4.2

F;16-238/Domain: trypsin homology <TRY>  
 F;1-122,42-58,136-201,168-182,191-220/Disulfide bonds: #status experimental  
 F;57,102,195/Active site: His, Asp, Ser #status experimental

Query Match 100.0%; Score 112; DB 1; Length 245;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPWQVSLQD 20  
 |||||:|||||:|||||:|||||  
 Db 16 IVNGEAVPGSWPWQVSLQD 35

## RESULT 2

KYBOB  
 chymotrypsin (EC 3.4.21.1) B precursor - bovine

N;Alternate names: chymotrypsinogen B  
 C;Species: Bos primigenius taurus (cattle)  
 C;Date: 08-Oct-1981 #sequence\_revision 08-Oct-1981 #text\_change 18-Jul-1997  
 C;Accession: A00953  
 R;Smillie, L.B.; Furka, A.; Nagabhushan, N.; Stevenson, K.J.; Parkes, C.O.  
 Nature 218, 343-346, 1968  
 A;Title: Structure of chymotrypsinogen B compared with chymotrypsinogen A and trypsinogen  
 A;Reference number: A00953; MUID:68238908; PMID:5649671  
 A;Accession: A00953

A;Molecule type: protein  
 A;Residues: 1-245 <SMI>  
 C;Comment: Chymotrypsinogen B is synthesized, along with chymotrypsinogen A, in the acinar cells of the pancreas of the bovine. The first activation cleavage, leading to pi-chymotrypsin B, occurs in the secretory granules.  
 C;Superfamily: trypsin; trypsin homology  
 C;Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen  
 F;1-15/Domain: propeptide #status experimental <PRO>  
 F;16-245/Product: chymotrypsin B #status experimental <MAT>  
 F;16-238/Domain: trypsin homology <TRY>  
 F;1-122,42-58,136-201,168-182,191-220/Disulfide bonds: #status experimental  
 F;57,102,195/Active site: His, Asp, Ser #status experimental

Query Match 97.3%; Score 109; DB 1; Length 245;

Best Local Similarity 95.0%; Pred. No. 9.7e-09;  
 Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPWQVSLQD 20  
 |||||:|||||:|||||:|||||  
 Db 16 IVNGEAVPGSWPWQVSLQD 35

## RESULT 3

A21195  
 chymotrypsin (EC 3.4.21.1) 2 precursor - dog

C;Species: Canis lupus familiaris (dog)  
 C;Date: 27-Jul-1990 #sequence\_revision 27-Jul-1990 #text\_change 22-Jun-1999  
 C;Accession: A21195  
 R;Pinaky, S.D.; LaForge, K.S.; Luc, V.; Scheele, G.  
 Proc. Natl. Acad. Sci. U.S.A. 80, 7486-7490, 1983  
 A;Title: Identification of cDNA clones encoding secretory isoenzyme forms: sequence data  
 A;Reference number: A21195; MUID:84170253; PMID:6584866  
 A;Accession: A21195

A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-263 <PIN>  
 A;Cross-references: GB:K01173; NID:g163945; PIDN:AAA30841.1; PID:g163946  
 C;Superfamily: trypsin; trypsin homology  
 C;Keywords: hydrolase; pancreas; protein digestion; serine proteinase  
 F;1-15/Domain: propeptide #status experimental <PRO>  
 F;16-245/Product: chymotrypsin B #status predicted <MAT>  
 F;1-122,42-58,136-201,168-182,191-220/Disulfide bonds: #status predicted

Query Match 97.3%; Score 109; DB 2; Length 263;

Best Local Similarity 95.0%; Pred. No. 1e-08;  
 Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPWQVSLQD 20  
 |||||:|||||:|||||:|||||  
 Db 34 IVNGEAVPGSWPWQVSLQD 53

## RESULT 4

A31299  
 chymotrypsin (EC 3.4.21.1) precursor - human

C;Species: Homo sapiens (man)  
 C;Date: 08-Jun-1989 #sequence\_revision 08-Jun-1989 #text\_change 22-Jun-1999  
 C;Accession: A31299  
 R;Tomita, N.; Izumoto, Y.; Horii, A.; Doi, S.; Yokouchi, H.; Ogawa, M.; Mori, T.; Matsubara, T.  
 Biochem. Biophys. Res. Commun. 158, 569-575, 1989  
 A;Title: Molecular cloning and nucleotide sequence of human pancreatic prechymotrypsinogen  
 A;Reference number: A31299; MUID:89134264; PMID:2917002  
 A;Accession: A31299

A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-263 <TOM>  
 A;Cross-references: GB:M24400; NID:g181189; PIDN:AAA52128.1; PID:g181190  
 C;Genetics:

A;Gene: GDB:CTRB1; CTRB  
 A;Cross-references: GDB:119820; OMIM:118890  
 A;Map position: 16q23.1-16q23.1  
 C;Superfamily: trypsin; trypsin homology  
 C;Keywords: hydrolase; pancreas; protein digestion; serine proteinase  
 F;1-15/Domain: propeptide #status predicted <PRO>  
 F;16-245/Product: chymotrypsin B #status predicted <MAT>  
 F;1-122,42-58,136-201,168-182,191-220/Disulfide bonds: #status predicted

Query Match 97.3%; Score 109; DB 2; Length 263;

Best Local Similarity 95.0%; Pred. No. 1e-08;  
 Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPWQVSLQD 20

|||||:|||||:|||||:|||||  
 Db 34 IVNGEAVPGSWPWQVSLQD 53

## RESULT 5

KYRTB

chymotrypsin (EC 3.4.21.1) B precursor - rat

N;Alternate names: chymotrypsinogen B  
 C;Species: Rattus norvegicus (Norway rat)  
 C;Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 18-Jun-1999  
 C;Accession: A22658  
 R;Bell, G.I.; Quinto, C.; Quiroga, M.; Valenzuela, P.; Craik, C.S.; Rutter, W.J.  
 J. Biol. Chem. 259, 14265-14270, 1984  
 A;Title: Isolation and sequence of a rat chymotrypsin B gene.  
 A;Reference number: A22658; MUID:85054881; PMID:6209274

A;Accession: A22658

A;Molecule type: DNA

A;Residues: 1-263 <BEL>

A;Cross-references: GB:K02298; NID:g203653; PIDN:AAA98732.1; PID:g203654

C;Genetics:

A;Introns: 18/1; 52/3; 79/2; 105/3; 166/1; 210/3

C;Superfamily: trypsin; trypsin homology

C;Keywords: hydrolase; pancreas; protein digestion; serine proteinase

F;1-18/Domain: signal sequence #status predicted <SIG>

F;19-33/Domain: propeptide #status predicted <PRO>

F;34-256/Product: chymotrypsin B #status predicted <MAT>

F;34-256/Domain: trypsin homology <TRY>

F;75,120,213/Active site: His, Asp, Ser #status predicted

Query Match 96.4%; Score 108; DB 1; Length 263;

Best Local Similarity 90.0%; Pred. No. 1.5e-08;  
 Matches 18; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPWQVSLQD 20

|||||:|||||:|||||:|||||  
 Db 34 IVNGEAVPGSWPWQVSLQD 53

## RESULT 6

B61333

chymotrypsin (EC 3.4.21.1) - edible frog (fragment)

C;Species: Rana esculenta (edible frog)

```

C>Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 07-May-1999
C:Accession: B61333
P:Pies, W.; Zwilling, R.; Woodbury, R.G.; Neurath, H.
FEBS Lett. 109, 45-49, 1980
A:Title: Amino-terminal amino acid sequences and the evolution of frog (Rana esculenta)
A:Reference number: A61333; MUID:80113255; PMID:6965480
A:Accession: B61333
A:Molecule type: protein
A:Residues: 1-20 <PI>
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; protein digestion; serine proteinase

Query Match          91.1%; Score 102; DB 2; Length 20;
Best Local Similarity 90.0%; Pred. No. 6.6e-09;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPQVSLQD 20
    ||||| ||||| ||||| |||||
DB 1 IVNGENAVPGSWPQVSLQN 20

RESULT 7
A23473
chymotrypsin-like proteinase (EC 3.4.21.-) - pig (tentative sequence) (fragments)
N:Alternate names: pancreatic elastase II [misidentification]
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 31-Mar-2000
C:Accession: A23473
R:Vered, M.; Gertler, A.; Burshtein, Y.
Int. J. Pept. Protein Res. 27, 183-190, 1986
A:Reference number: A23473; MUID:86194934; PMID:3634756
A:Accession: A23473
A:Molecule type: protein
A:Residues: 1-126 <VER>
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase

Query Match          87.5%; Score 98; DB 2; Length 126;
Best Local Similarity 70.0%; Pred. No. 2e-07;
Matches 14; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPQVSLQD 20
    ||||| ||||| ||||| |||||
DB 12 IVBGZBAVPGSWPQVSLZB 31

RESULT 8
S72219
chymotrypsin B - Atlantic cod (fragments)
C:Species: Gadus morhua (Atlantic cod)
C>Date: 19-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 07-Aug-1998
C:Accession: S72219
R:Leth-Larsen, R.; Ageeissoren, B.; Thorolfsson, M.; Norregaard-Madsen, M.; Hojrup, P.
Biochim. Biophys. Acta 1997, 49-56, 1996
A:Title: Structure of chymotrypsin variant B from Atlantic cod, Gadus morhua.
A:Reference number: S72219; MUID:96439045; PMID:8841380
A:Accession: S72219
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-14;15-244 <LET>
C:Superfamily: trypsin; trypsin homology
F:15-237/Domain: trypsin homology <TRY>

```

138136  
chymotrypsin-like proteinase (EC 3.4.21.-) CTRL-1 - human  
C:Species: Homo sapiens (man)  
C:Date: 06-Sep-1996 #sequence\_revision 06-Sep-1996 #text\_change 22-Jun-1999  
C:Accession: I38136  
R:Larsen, F.; Solhelt, J.; Kristensen, T.; Kolsto, A.B.; Prydz, H.  
Hum. Mol. Genet. 2, 1589-1595, 1993  
A:Title: A tight cluster of five unrelated human genes on chromosome 16q22.1.  
A:Reference number: I38133; MUID:94093544; PMID:8268911  
A:Accession: I38136  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-264 <RES>  
A:Cross-references: EMBL:X71874; NID:g406226; PIDN:CAA50710.1; PID:g406228  
C:Genetics:  
A:Gene: GDB:CTRL  
A:Cross-references: GDB:204061  
A:Map position: 16q22.1-16q22.1  
A:Introns: 18/3; 52/3; 79/2; 106/3; 167/1; 211/3  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase  
C:Keywords: hydrolase; serine proteinase  
F:34-257/Domain: trypsin homology <TRY>  
F:75,121,214/Active site: His, Asp, Ser #status predicted

Query Match 86.6%; Score 97; DB 2; Length 264;  
Best Local Similarity 90.0%; Pred. No. 6.3e-07;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps

QY 1 IVNGEEAVPGSWPQVSLQD 20  
||||| |||||||  
Db 34 IVNGENAVLGSPWQVSLQD 53

RESULT 10  
S47537  
chymotrypsin (EC 3.4.21.1) precursor - Atlantic cod  
C:Species: Gadus morhua (Atlantic cod)  
C:Date: 26-Dec-1994 #sequence\_revision 03-Aug-1995 #text\_change 22-Jun-1999  
C:Accession: S47537; S43163  
R:Gudmundsdottir, A.; Oskarsson, S.; Eakin, A.E.; Craik, C.S.; Bjarnason, J.B.  
Biochim. Biophys. Acta 1219, 211-214, 1994  
A:Title: Atlantic cod cDNA encoding a psychrophilic chymotrypsinogen.  
A:Reference number: S47537; MUID:94368860; PMID:8086467  
A:Accession: S47537  
A:Molecule type: mRNA  
A:Residues: 1-263 <GUD>  
A:Cross-references: EMBL:X78490; NID:g468750; PIDN:CAA55242.1; PID:g468751  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; protein digestion; serine proteinase  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:19-263/Product: chymotrypsin #status predicted <MAT>  
F:34-256/Domain: trypsin homology <TRY>  
F:75,120,213/Active site: His, Asp, Ser #status predicted

Query Match 85.7%; Score 96; DB 2; Length 263;  
Best Local Similarity 90.0%; Pred. No. 8.9e-07;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps

QY 1 IVNGEEAVPGSWPQVSLQD 20  
||||| |||||||  
Db 34 IVNGENAVPHSHWQVSLQD 53

RESULT 11  
B33257  
pancreatic elastase (EC 3.4.21.36) III - pig (fragment)  
N:Alternate names: proteinase E  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 21-Feb-1990 #sequence\_revision 21-Feb-1990 #text\_change 30-Jun-1995  
C:Accession: B33257  
R:Aviles, F.X.; Pascual, R.; Salva, M.; Bonicel, J.; Puigserver, A.  
Biochim. Biophys. Res. Commun. 163, 1191-1196, 1989  
A:Title: Generation of a subunit III-like protein by autolysis of human and pig

A;Reference number: A33257; MUID:89392022; PMID:2675835

A;Accession: B33257

A;Status: preliminary

A;Molecule type: protein

A;Residues: 1-31 <AVI>

C;Superfamily: trypsin; trypsin homology

C;Keywords: hydrolase; serine proteinase

Query Match 83.0%; Score 93; DB 2; Length 31;

Best Local Similarity 84.2%; Pred. No. 2.3e-07;

Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPQVSLQ 19

:||||:|||||

Db 12 VVNGEDAVPYSPWPQVSLQ 30

RESULT 12

B29934

pancreatic elastase (EC 3.4.21.36) IIIB precursor - human

N;Alternate names: 35K glycoprotein; pancreatic protein P35; protein G32; proteinase E

C;Species: Homo sapiens (man)

C;Date: 15-Dec-1988 #sequence\_revision 15-Dec-1988 #text\_change 22-Jun-1999

C;Accession: B29934; A90516; A27206; A33257; A28932; S04599; S04490

R;Tani, T.; Ohsumi, J.; Mita, K.; Takiguchi, Y.

J. Biol. Chem. 263, 1231-1239, 1988

A;Title: Identification of a novel class of elastase isozyme, human pancreatic elastase

A;Reference number: A92664; MUID:88087253; PMID:2826474

A;Accession: B29934

A;Molecule type: mRNA

A;Residues: 1-270 <TAN>

A;Cross-references: GB:M18692; NID:G607029; PIDN:AAA58454.1; PID:G182035

R;Shen, W.; Fletcher, T.S.; Largman, C.

Biochemistry 26, 3447-3452, 1987

A;Title: Primary structure of human pancreatic protease E determined by sequence analysis

A;Reference number: A90516; MUID:88000545; PMID:3477287

A;Accession: A90516

A;Molecule type: mRNA

A;Residues: G, 5-63, 'G', 65-78, 'W', 80-118, 'G', 120-163, 'P', 165-270 <SHE>

R;Fletcher, T.S.

submitted to GenBank, August 1987

A;Reference number: A94507

A;Contents: revision to residue 119

A;Accession: A27206

A;Molecule type: DNA

A;Residues: 'G', 5-63, 'G', 65-78, 'W', 80-163, 'P', 165-270 <FLE>

R;Aviles, F.X.; Pascual, R.; Salva, M.; Bonicel, J.; Puigserver, A.

Biochem. Biophys. Res. Commun. 163, 1191-1196, 1989

A;Title: Characterization of a subunit III-like protein by autolysis of human and porcine pro-

A;Reference number: A33257; MUID:89392022; PMID:2675835

A;Accession: A33257

A;Molecule type: protein

A;Residues: 18-57 <AVI>

R;Guy-Crotte, O.; Barthe, C.; Basso, D.; Fournet, B.; Figarella, C.

Biochem. Biophys. Res. Commun. 156, 318-322, 1989

A;Title: Characterization of two glycoproteins of human pancreatic juice: P35, a truncated

A;Reference number: A28932; MUID:89025862; PMID:3178837

A;Accession: A28932

A;Molecule type: protein

A;Residues: 'X', 32-52, 'X', 54-55, 'XXX', 59-63 <GUY>

R;Mouillard, M.; Kerfelec, B.; Mallet, B.; Chapus, C.

FEBS Lett. 250, 166-170, 1989

A;Title: Identification of a procarboxypeptidase A-truncated protease E binary complex i

A;Reference number: S04999; MUID:89325560; PMID:2753124

A;Accession: S04999

A;Molecule type: protein

A;Residues: 31-50 <MOU>

R;Wendorf, P.; Geyer, R.; Sziegoleit, A.; Linder, D.

FEBS Lett. 249, 275-278, 1989

A;Title: Localization and characterization of the glycosylation site of human pancreatic

A;Reference number: S04490; MUID:89289996; PMID:2737288

A;Accession: S04490

A;Molecule type: protein

A;Residues: 94-128;132-164 <WEN>

C;Superfamily: trypsin; trypsin homology

C;Keywords: glycoprotein; hydrolase; serine proteinase; zymogen

F;1-17/Domain: signal sequence #status predicted <SIG>

F;18-28/Domain: activation peptide #status predicted <ACT>

F;29-270/Product: (or 31-270) pancreatic elastase IIIB #status predicted <MAT>

F;73,123,217/Active site: His, Asp, Ser #status predicted

F;114/Binding site: carbohydrate (Asn) (covalent) #status experimental

F;153/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 83.0%; Score 93; DB 2; Length 270;

Best Local Similarity 84.2%; Pred. No. 2.5e-06;

Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPQVSLQ 19

:||||:|||||

Db 29 VVNGEDAVPYSPWPQVSLQ 47

RESULT 13

A29934

pancreatic elastase (EC 3.4.21.36) IIIB precursor - human

N;Alternate names: protease E

C;Species: Homo sapiens (man)

C;Date: 15-Dec-1988 #sequence\_revision 15-Dec-1988 #text\_change 20-Jun-2000

C;Accession: A29934; JX0045

R;Tani, T.; Ohsumi, J.; Mita, K.; Takiguchi, Y.

J. Biol. Chem. 263, 1231-1239, 1988

A;Title: Identification of a novel class of elastase isozyme, human pancreatic elastase I

A;Reference number: A92664; MUID:88087253; PMID:2826474

A;Accession: A29934

A;Molecule type: DNA

A;Residues: 1-270 <TAN>

A;Cross-references: GB:J03516

R;Shirao, Y.; Takemura, K.; Yoshida, H.; Sato, Y.; Iijima, H.; Shimada, Y.; Mikayama, T.

J. Biochem. 104, 259-264, 1988

A;Title: Molecular cloning of complementary DNA encoding one of the human pancreatic pro-

A;Reference number: JX0045; MUID:89034017; PMID:2460440

A;Accession: JX0045

A;Molecule type: mRNA

A;Residues: 1-240, 'G', 242-270 <SHI>

A;Cross-references: GB:D00306; NID:G220013; PIDN:BA00212.1; PID:G220014

C;Comment: This enzyme is an alanine-specific serine proteinase that has little elastolyt

C;Genetics: 15/1; 43/3; 76/2; 121/2; 167/1; 214/3; 265/3

C;Superfamily: trypsin; trypsin homology

C;Keywords: hydrolase; pancreas; serine proteinase

F;1-15/Domain: signal sequence #status predicted <SIG>

F;16-28/Domain: propeptide #status predicted <PRO>

F;29-270/Product: pancreatic elastase IIIA #status experimental <MAT>

F;73,123,217/Active site: His, Asp, Ser #status predicted

Query Match 78.6%; Score 88; DB 2; Length 270;

Best Local Similarity 78.9%; Pred. No. 1.4e-05;

Matches 15; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPQVSLQ 19

:||||:|||||

Db 29 VVNGEDAVPYSPWPQVSLQ 47

RESULT 14

PU0036

serine proteinase (EC 3.4.21.-) - bovine (fragment)

N;Alternate names: trypsin-like proteinase bPTLP

C;Species: Bos primigenius taurus (cattle)

C;Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 22-Apr-1995

C;Accession: PU0036; PU0039

R;Tsuji, A.; Sakiyama, K.; Edazawa, K.; Nagata, K.; Sasaki, Y.; Nagamune, H.; Matsuda, Y.

submitted to JIPID, September 1994

A;Description: Purification and characterization of a novel serine proteinase from bovine

```

A:Reference number: PU0036
A:Accession: PU0036
A:Molecule type: Protein
A:Residues: 1-23 <ISU>
A:Experimental source: pancreas
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase

Query Match          76.8%; Score 86; DB 2; Length 23;
Best Local Similarity 68.4%; Pred. No. 1.8e-06;
Matches 13; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVNGEEAVPGSWPWQVSLQ 19
      :| |||:| |||||:|
Db 1 VVGGEAIPHSWPWQISLQ 19

RESULT 15
B26823
pancreatic elastase II (RC 3.4.21.71) A precursor - human
C:Species: Homo sapiens (man)
C:Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 20-Jun-2000
C:Accession: B26823; A27431; S34491
R:Kawashima, I.; Tani, T.; Shimoda, K.; Takiguchi, Y.
DNA 6, 163-172, 1987
A:Title: Characterization of pancreatic elastase II cDNAs: two elastase II mRNAs are exp
A:Reference number: A90958; MUID:87217962; PMID:3646943
A:Accession: B26823
A:Molecule type: mRNA
A:Residues: 1-269 <KW>
A:Cross-references: GB:M16652; NID:g182057; PIDN:AAA52380.1; PID:g182058
R:Fletcher, T.S.; Shen, W.F.; Largman, C.
Biochemistry 26, 7256-7261, 1987
A:Title: Primary structure of human pancreatic elastase 2 determined by sequence analysis
A:Reference number: A27432; MUID:88107669; PMID:3427074
A:Accession: A27432
A:Molecule type: mRNA
A:Residues: 1-269 <FLE>
A:Cross-references: GB:M16631; NID:g182022; PIDN:AAA52374.1; PID:g182023
R:Shiraga, Y.; Yoshida, H.; Matsuki, S.; Takemura, K.; Ikeda, N.; Shimada, Y.; Ozawa, T.
J. Biochem. 102, 1555-1563, 1987
A:Title: Molecular cloning and expression in Escherichia coli of a cDNA encoding human p
A:Reference number: A41431; MUID:88198076; PMID:2834346
A:Accession: A41431
A:Molecule type: mRNA
A:Residues: 1-201, 'V', 203-269 <SHI>
A:Cross-references: GB:D00236; NID:g219619; PIDN:BA00165.1; PID:g219620
R:Moulard, M.; Michon, T.; Kerfelec, B.; Chapus, C.
FEBS Lett. 261, 179-183, 1990
A:Title: Further studies on the human pancreatic binary complexes involving procarboxype
A:Reference number: S08253; MUID:90169111; PMID:2307232
A:Accession: S34491
A:Molecule type: protein
A:Residues: 'X', 18-50 <MOU>
C:Genetics:
A:Gene: GDB:ELAI
A:Cross-references: GDB:119866; OMIM:130120
A:Map position: 12pter-12qter
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; serine proteinase
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-28/Domain: propeptide #status predicted <PRO>
F:29-269/Product: pancreatic elastase IIA #status predicted <MAT>
F:29-262/Domain: trypsin homology <TRY>
F:73,121,216/Active site: His, Asp, Ser #status predicted

Query Match          76.8%; Score 86; DB 2; Length 269;
Best Local Similarity 78.9%; Pred. No. 2.8e-05;
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEEAVPGSWPWQVSLQ 19
      :| |||:| |||||:|

```

Db 29 VVGGEARPNPNSWPWQVSLQ 47

Search completed: February 12, 2003, 10:28:52  
Job time : 6.01493 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:04:45 ; Search time 2.68657 Seconds  
(without alignments)  
308.768 Million cell updates/sec

Title: US-10-036-371-6

Perfect score: 112

Sequence: 1 IVNGEAVPGSWPWQVSLQD 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	112	100.0	245	1	CTRA_BOVIN	P00766	bos taurus
2	109	97.3	245	1	CTRB_BOVIN	P00767	bos taurus
3	108	97.3	263	1	CTR2_CANFA	P04813	canis famil
4	109	97.3	263	1	CTRB_HUMAN	P17538	homo sapien
5	108	96.4	263	1	CTRB_RAT	P07338	rattus norv
6	98	87.5	245	1	CTRB_GADMO	P06646	gadus morhu
7	97	86.6	264	1	CTRL_HUMAN	P40313	homo sapien
8	96	85.7	263	1	CTRA_GADMO	P47796	gadus morhu
9	93	83.0	270	1	EL3B_HUMAN	P08661	homo sapien
10	90	80.4	268	1	CLCR_RAT	P55091	rattus norv
11	88	78.6	270	1	EL3A_HUMAN	P09093	homo sapien
12	86	76.8	269	1	EL2A_HUMAN	P08217	homo sapien
13	85	75.9	253	1	CAC3_BOVIN	P05805	bos taurus
14	84	75.0	271	1	EL2_RAT	P00774	rattus norv
15	84	75.0	492	1	TMS2_HUMAN	O15393	homo sapien
16	83	74.1	269	1	EL2_BOVIN	O29461	bos taurus
17	83	74.1	269	1	EL2_MOUSE	P08419	sus scrofa
18	83	74.1	271	1	EL2_MOUSE	P05208	mus musculus
19	81	72.3	269	1	EL2B_HUMAN	P08218	homo sapien
20	80	71.4	268	1	CLCR_HUMAN	P09895	homo sapien
21	75	67.0	60	1	ACRO_CAPHI	P10626	capra hircu
22	75	67.0	437	1	TMS4_HUMAN	Q9nr84	homo sapien
23	73	65.2	415	1	ACRO_PIG	P08001	sus scrofa
24	73	65.2	418	1	HATT_HUMAN	O60235	homo sapien
25	72	64.3	490	1	TMS2_MOUSE	Q9jiq8	mus musculus
26	69	61.6	20	1	COG1_PACRM	P20731	paralithode
27	69	61.6	436	1	ACRO_MOUSE	P23578	mus musculus
28	69	61.6	638	1	KAL_MOUSE	P26262	mus musculus
29	68	60.7	20	1	ELAS_GADMO	P32197	gadus morhu
30	68	60.7	235	1	TRYD_HUMAN	Q9bzj3	homo sapien
31	68	60.7	270	1	TRYT_MERUN	P50342	meriones un
32	68	60.7	273	1	MCT7_RAT	P27435	rattus norv
33	68	60.7	275	1	TRB1_HUMAN	Q15661	homo sapien

34	68	60.7	275	1	TRB2_HUMAN	P20231	homo sapien
35	68	60.7	275	1	TRYA_HUMAN	P15157	homo sapien
36	68	60.7	422	1	DESI_HUMAN	Q9u152	homo sapien
37	68	60.7	431	1	ACRO_RABIT	P48038	oryctolagus
38	67	59.8	273	1	MCT7_MOUSE	Q02844	mus musculus
39	67	59.8	274	1	MCT6_RAT	P50343	rattus norv
40	67	59.8	276	1	MCT6_MOUSE	P21845	mus musculus
41	67	59.8	290	1	WPN_HUMAN	Q9bq13	homo sapien
42	67	59.8	342	1	PSS8_MOUSE	Q9esd1	mus musculus
43	67	59.8	342	1	PSS8_RAT	Q9es87	rattus norv
44	67	59.8	343	1	PLMN_SHEEP	P81286	ovis aries
45	67	59.8	343	1	PSS8_HUMAN	Q16651	homo sapien

## ALIGNMENTS

RESULT 1  
CTRA\_BOVIN STANDARD; PRT; 245 AA.  
AC P00766;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Chymotrypsinogen A [EC 3.4.21.1].  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE, DISULFIDE BONDS, AND ACTIVE SITE.  
RX MEDLINE=67181721; PubMed=5971783;  
RA Brown J.R., Hartley B.S.;  
RT "Location of disulphide bridges by diagonal paper electrophoresis.  
RL The disulphide bridges of bovine chymotrypsinogen A.";  
RN [2]  
RP REVISION TO 102.  
RX MEDLINE=69106266; PubMed=5764436;  
RA Blow D.M., Birktoft J.J., Hartley B.S.;  
RT "Role of a buried acid group in the mechanism of action of  
RL chymotrypsin.";  
RN [3]  
RP Nature 221:337-340(1969).  
RA Hartley B.S.;  
RT "Amino-acid sequence of bovine chymotrypsinogen-A.";  
RN [4]  
RP Nature 201:1284-1287(1964).  
RX SEQUENCE, AND DISULFIDE BONDS.  
RA Meloun B., Klubb I., Kostka V., Moravsek L., Prusik Z., Vanacek J.,  
RT "Histidine residues in the active centres of some 'serine'  
RL proteinases.";  
RN [5]  
RP Biochem. J. 101:232-241(1966).  
RX MEDLINE=67181723; PubMed=5971785;  
RA Smillie L.B., Hartley B.S.;  
RT "Histidine residues in the active centres of some 'serine'  
RL proteinases.";  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY.  
RX MEDLINE=72035052; PubMed=4399050;  
RA Birktoft J.J., Blow D.M., Henderson R., Steitz T.A.;  
RT "I. Serine proteinases. The structure of alpha-chymotrypsin.";  
RN [7]  
RP Philos. Trans. R. Soc. Lond., B, Biol. Sci. 257:67-76(1970).  
RX X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF CHYMOTRYPSINOGEN.  
RN [7]  
RP Freer S.T., Kraut J., Robertus J.D., Wright H.T., Xuong N.H.;

RT "Chymotrypsinogen: 2.5-A crystal structure, comparison with alpha-chymotrypsin, and implications for zymogen activation.";  
RL Biochemistry 9:1997-2009(1970).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF GAMMA-CHYMOTRYPSIN.  
RX MEDLINE=82078042; PubMed=6914398;  
RA Cohen G.H., Silverton E.W., Davies D.R.;  
RT "Refined crystal structure of gamma-chymotrypsin at 1.9-A resolution.  
RT Comparison with other pancreatic serine proteases.";  
RL J. Mol. Biol. 148:449-479(1981).  
RN [9]  
RP X-RAY CRYSTALLOGRAPHY (1.68 ANGSTROMS) OF ALPHA-CHYMOTRYPSIN.  
RX MEDLINE=86011575; PubMed=4046030;  
RA Tsukada H., Blow D.M.;  
RT "Structure of alpha-chymotrypsin refined at 1.68-A resolution.";  
RL J. Mol. Biol. 184:703-711(1985).  
CC -|- CATALYTIC ACTIVITY: Preferential cleavage: Tyr|-Xaa, Trp|-Xaa,  
Phe|-Xaa, Leu|-Xaa.  
CC -|- SUBCELLULAR LOCATION: Extracellular.  
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
CC -|- DATABASE: NAME=Worthington enzyme manual;  
WWW="http://www.worthington-biochem.com/manual/C/CHY.html".  
DR PIR; A09552; KYBOA.  
DR PDB; 2CGA; 15-APR-90.  
DR PDB; 2CHA; 31-MAY-84.  
DR PDB; 4CHA; 29-OCT-85.  
DR PDB; 5CHA; 16-OCT-87.  
DR PDB; 6CHA; 16-OCT-87.  
DR PDB; 1CHG; 27-JAN-84.  
DR PDB; 1CHO; 16-JUL-88.  
DR PDB; 2GCH; 31-MAY-84.  
DR PDB; 3GCH; 15-OCT-92.  
DR PDB; 4GCH; 15-OCT-90.  
DR PDB; 5GCH; 15-OCT-90.  
DR PDB; 6GCH; 15-OCT-90.  
DR PDB; 7GCH; 15-OCT-90.  
DR PDB; 8GCH; 15-JUL-93.  
DR PDB; 1GCT; 15-OCT-91.  
DR PDB; 2GCT; 15-OCT-91.  
DR PDB; 3GCT; 15-OCT-91.  
DR PDB; 1ACB; 31-OCT-93.  
DR PDB; 1GMC; 31-OCT-93.  
DR PDB; 1GMD; 31-OCT-93.  
DR PDB; 1CGI; 30-APR-94.  
DR PDB; 1CGJ; 30-APR-94.  
DR PDB; 1CGK; 30-APR-94.  
DR PDB; 1GHA; 22-JUN-94.  
DR PDB; 1GHB; 22-JUN-94.  
DR PDB; 1GMH; 30-SEP-94.  
DR PDB; 2GMT; 01-NOV-94.  
DR PDB; 1NTN; 17-AUG-96.  
DR PDB; 1AB9; 20-AUG-97.  
DR PDB; 1AFQ; 17-SEP-97.  
DR PDB; 1CA0; 23-JUL-97.  
DR PDB; 1CBW; 23-JUL-97.  
DR PDB; 1VGC; 12-NOV-97.  
DR PDB; 2VGC; 12-NOV-97.  
DR PDB; 3VGC; 12-NOV-97.  
DR PDB; 4VGC; 12-NOV-97.  
DR PDB; 1HJA; 14-JAN-98.  
DR MEROPS; S01.001; --  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen;  
3D-structure.  
FT CHAIN 1 13 CHYMOTRYPSIN A, A CHAIN.  
FT CHAIN 16 146 CHYMOTRYPSIN A, B CHAIN.

FT CHAIN 149 245 CHYMOTRYPSIN A, C CHAIN.  
FT ACT\_SITE 57 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 102 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 195 CHARGE RELAY SYSTEM.  
FT DISULFID 1 122  
FT DISULFID 42 58  
FT DISULFID 136 201  
FT DISULFID 168 182  
FT DISULFID 191 220  
FT HELIX 12 15  
FT STRAND 20 21  
FT TURN 24 25  
FT TURN 28 29  
FT STRAND 30 34  
FT STRAND 40 46  
FT STRAND 51 54  
FT TURN 57 58  
FT TURN 62 63  
FT STRAND 65 68  
FT TURN 69 69  
FT STRAND 72 72  
FT STRAND 73 74  
FT STRAND 81 90  
FT TURN 92 93  
FT STRAND 95 95  
FT TURN 96 99  
FT STRAND 100 100  
FT TURN 101 101  
FT STRAND 104 108  
FT STRAND 122 122  
FT TURN 126 127  
FT TURN 132 133  
FT STRAND 135 140  
FT TURN 146 148  
FT STRAND 154 154  
FT STRAND 156 163  
FT HELIX 165 168  
FT TURN 169 172  
FT HELIX 173 175  
FT STRAND 180 184  
FT TURN 192 193  
FT TURN 195 196  
FT STRAND 198 203  
FT TURN 204 205  
FT STRAND 206 216  
FT TURN 218 219  
FT TURN 222 223  
FT STRAND 225 230  
FT HELIX 231 233  
FT TURN 234 234  
FT HELIX 235 243  
SQ SEQUENCE 245 AA; 25666 MW; 91A9F28E2F3E3142 CRC64;  
Query Match 100.0%; Score 112; DB 1; Length 245;  
Best Local Similarity 100.0%; Pred. No. 4.1e-09; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0;  
QY 1 IVNGEEAVPGSWPQVSLQD 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 16 IVNGEEAVPGSWPQVSLQD 35  
| | | | | | | | | | | | | | | | | | | | | |  
RESULT 2  
CTRB BOVIN STANDARD; PRT; 245 AA.  
ID -CTRB BOVIN  
AC P00767;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Chymotrypsinogen B (EC 3.4.21.1).  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;



```

CC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE, DISULFIDE BONDS, AND ACTIVE SITE.
RX MEDLINE=68238908; PubMed=5649671;
RA Smilie L.B., Furka A., Nagabhushan N., Stevenson K.J., Parkes C.O.;
RT "Structure of chymotrypsinogen B compared with chymotrypsinogen A and
trypsinogen.";
RL Nature 218:343-346(1968).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
Phe-|-Xaa, Leu-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- DATABASE: NAME=Washington enzyme manual;
WWW="http://www.worthington-biochem.com/manual/C/CHV.html".
DR PIR; A00953; KYBOB.
DR HSP; P00766; IACB.
DR MEROPS; S01.152; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal.
FT CHAIN 1 13 CHYMOTRYPSIN B, A CHAIN.
FT CHAIN 16 146 CHYMOTRYPSIN B, B CHAIN.
FT CHAIN 149 245 CHYMOTRYPSIN B, C CHAIN.
FT ACT_SITE 57 57 CHARGE RELAY SYSTEM.
FT ACT_SITE 102 102 CHARGE RELAY SYSTEM.
FT ACT_SITE 195 195 CHARGE RELAY SYSTEM.
FT DISULFID 1 122
FT DISULFID 42 58
FT DISULFID 136 201
FT DISULFID 168 182
FT DISULFID 191 220
SQ SEQUENCE 245 AA; 25755 MW; 678016446FF5FB5 CRC64;

Query Match 97.3%; Score 109; DB 1; Length 245;
Best Local Similarity 95.0%; Pred. No. 1.1e-08;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPQVSLQD 20
Db 16 IVNGEDAVPGSWPQVSLQD 35
|||||:|||||:|||||:|||||:|||||:
RESULT 3
ID CTR2_CANFA STANDARD; PRT; 263 AA.
AC P04813;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymotrypsinogen 2 precursor (EC 3.4.21.1).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84170253; PubMed=6584866;
RA Pinsky S.D., Laforge K.S., Luc V., Scheele G.;
RT "Identification of cDNA clones encoding secretory isoenzyme forms:
sequence determination of canine pancreatic chymotrypsinogen 2
mRNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:7486-7490(1983).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
Phe-|-Xaa, Leu-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

```

---

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K01173; AAA30841.1; -.
DR PIR; A21195; A21195.
DR HSP; P00766; IACB.
DR MEROPS; S01.152; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal.
FT SIGNAL 1 18
FT CHAIN 19 263 CHYMOTRYPSINOGEN 2.
FT CHAIN 19 31 CHYMOTRYPSIN 2, A CHAIN.
FT CHAIN 34 164 CHYMOTRYPSIN 2, B CHAIN.
FT CHAIN 167 263 CHYMOTRYPSIN 2, C CHAIN.
FT ACT_SITE 75 75 CHARGE RELAY SYSTEM.
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM.
FT ACT_SITE 213 213 CHARGE RELAY SYSTEM.
FT DISULFID 19 140 BY SIMILARITY.
FT DISULFID 60 76 BY SIMILARITY.
FT DISULFID 154 219 BY SIMILARITY.
FT DISULFID 186 200 BY SIMILARITY.
FT DISULFID 209 238 BY SIMILARITY.
SQ SEQUENCE 263 AA; 27787 MW; 2A2F449D813B3961 CRC64;

Query Match 97.3%; Score 109; DB 1; Length 263;
Best Local Similarity 95.0%; Pred. No. 1.2e-08;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPQVSLQD 20
Db 34 IVNGEDAVPGSWPQVSLQD 53
|||||:|||||:|||||:|||||:|||||:
RESULT 4
ID CTRB_HUMAN STANDARD; PRT; 263 AA.
AC P17538;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymotrypsinogen B precursor (EC 3.4.21.1).
GN CTRB1 OR CTRB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89134284; PubMed=2917002;
RA Tomita N., Izumoto Y., Horii A., Doi S., Yokouchi H., Ogawa M.,
Mori T., Matsubara K.;
RT "Molecular cloning and nucleotide sequence of human pancreatic
prechymotrypsinogen cDNA.";
RL Biochem. Biophys. Res. Commun. 158:569-575(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

```

```

CC -|- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
CC Phe-|-Xaa, Leu-|-Xaa.
CC -|- SUBCELLULAR LOCATION: Extracellular.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24400; AAA52128.1; -.
CC DR EMBL; BC005385; AAH05385.1; -.
CC DR PIR; A31299; A31299.
CC DR HSP; P00766; ICHG.
CC DR MEROPS; S01.152; -.
CC DR Genew; HGNC:2521; CTRB1.
CC DR MIM; 118890; -.
CC DR InterPro; IPR001314; Chymotrypsin.
CC DR InterPro; IPR001254; Ser_protease_Try.
CC DR Pfam; PF00089; trypsin; 1.
CC DR PRINTS; PR00722; CHYMOTRYPSIN.
CC DR SMART; SM00020; Tryp_Spc; 1.
CC DR PROSITE; PS00240; TRYPSIN_DOM; 1.
CC DR PROSITE; PS00134; TRYPSIN_HIS; 1.
CC DR PROSITE; PS00135; TRYPSIN_SER; 1.
CC DR PROSITE; PS00135; TRYPSIN_SER; 1.
CC KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal.
CC FT SIGNAL 1 18
CC FT CHAIN 19 263 CHYMOTRYPSINOGEN B.
CC FT CHAIN 19 31 CHYMOTRYPSIN B, A CHAIN.
CC FT CHAIN 34 164 CHYMOTRYPSIN B, B CHAIN.
CC FT CHAIN 167 263 CHYMOTRYPSIN B, C CHAIN.
CC FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT DISULFID 19 140 BY SIMILARITY.
CC FT DISULFID 60 76 BY SIMILARITY.
CC FT DISULFID 154 219 BY SIMILARITY.
CC FT DISULFID 186 200 BY SIMILARITY.
CC FT DISULFID 209 238 BY SIMILARITY.
CC FT DISULFID 209 238 BY SIMILARITY.
CC SQ SEQUENCE 263 AA; 27870 MW; 4C1C055A490B8701 CRC64;

Query Match 97.3%; Score 109; DB 1; Length 263;
Best Local Similarity 95.0%; Pred. No. 1.2e-08;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPQVSLQD 20
Db 34 IVNGEAVPGSWPQVSLQD 53

RESULT 5
CTRB_RAT
ID_CTRB_RAT STANDARD; PRT; 263 AA.
AC P07336;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymotrypsinogen B precursor (EC 3.4.21.1).
GN CTRB1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85054881; PubMed=6209274;
RA Bell G.I.; Quinto C.; Quiroga M.; Valenzuela P.; Craik C.S.;
RA Rutter W.J.;
RT "Isolation and sequence of a rat chymotrypsin B gene.";
RL J. Biol. Chem. 259:14265-14270(1984).

```

```

CC -|- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
CC Phe-|-Xaa, Leu-|-Xaa.
CC -|- SUBCELLULAR LOCATION: Extracellular.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K02298; AAA98732.1; -.
CC DR EMBL; A22658; KYRTB.
CC DR HSP; P00766; ICHG.
CC DR MEROPS; S01.152; -.
CC DR InterPro; IPR001314; Chymotrypsin.
CC DR InterPro; IPR001254; Ser_protease_Try.
CC DR Pfam; PF00089; trypsin; 1.
CC DR PRINTS; PR00722; CHYMOTRYPSIN.
CC DR SMART; SM00020; Tryp_Spc; 1.
CC DR PROSITE; PS00240; TRYPSIN_DOM; 1.
CC DR PROSITE; PS00134; TRYPSIN_HIS; 1.
CC DR PROSITE; PS00135; TRYPSIN_SER; 1.
CC KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal.
CC FT SIGNAL 1 18
CC FT CHAIN 19 263 CHYMOTRYPSINOGEN B.
CC FT CHAIN 19 31 CHYMOTRYPSIN B, A CHAIN.
CC FT CHAIN 34 164 CHYMOTRYPSIN B, B CHAIN.
CC FT CHAIN 167 263 CHYMOTRYPSIN B, C CHAIN.
CC FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT DISULFID 19 140 BY SIMILARITY.
CC FT DISULFID 60 76 BY SIMILARITY.
CC FT DISULFID 154 219 BY SIMILARITY.
CC FT DISULFID 186 200 BY SIMILARITY.
CC FT DISULFID 209 238 BY SIMILARITY.
CC FT DISULFID 209 238 BY SIMILARITY.
CC SQ SEQUENCE 263 AA; 27849 MW; ACAFD8ACF8C4DA6D CRC64;

Query Match 96.4%; Score 108; DB 1; Length 263;
Best Local Similarity 90.0%; Pred. No. 1.6e-08;
Matches 18; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPQVSLQD 20
Db 34 IVNGEAVPGSWPQVSLQD 53

RESULT 6
CTRB_GADMO
ID_CTRB_GADMO STANDARD; PRT; 245 AA.
AC P80646;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymotrypsin B (EC 3.4.21.1).
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.
OX NCBI_TaxID=8049;
RN [1]
RP SEQUENCE.
RC TISSUE=Pyloric caeca;
RX MEDLINE=96439045; PubMed=8841380;
RA Leth-Larsen R.; Aegirsson B.; Thorolfsson M.; Noerregaard-Madsen M.;
RA Hoejrup P.;
RT "Structure of chymotrypsin variant B from Atlantic cod, Gadus
RT morhua.";
RL Biochim. Biophys. Acta 1297:49-56(1996).
RN [2]

```



CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X78490; CAA35242.1; -.  
DR HSSP; P00766; 1CHG.  
DR MEROPS; S01.152; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS50240; TRYPsin DOM; 1.  
DR PROSITE; PS00134; TRYPsin HIS; 1.  
DR PROSITE; PS00135; TRYPsin SER; 1.  
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal.  
FT SIGNAL 1 18  
FT CHAIN 19 263 CHYMOTRYPSIN A.  
FT ACT\_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 19 140 BY SIMILARITY.  
FT DISULFID 60 76 BY SIMILARITY.  
FT DISULFID 154 219 BY SIMILARITY.  
FT DISULFID 186 200 BY SIMILARITY.  
FT DISULFID 209 238 BY SIMILARITY.  
FT CONFLICT 21 21 R -> S (IN REF. 2).  
FT CONFLICT 25 25 S -> Q (IN REF. 2).  
FT CONFLICT 29 29 T -> S (IN REF. 2).  
FT CONFLICT 44 44 S -> T (IN REF. 2).  
FT CONFLICT 46 46 S -> Y (IN REF. 2).  
SQ SEQUENCE 263 AA; 28294 MW; 47AAC699A0A64PBB CRC64;  
  
Query Match 85.7%; Score 96; DB 1; Length 263;  
Best Local Similarity 90.0%; Pred. No. 8.6e-07;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 IVNGEEAVPGSWPQVSLQD 20  
DQ 34 IVNGEEAVPHSWQVSLQD 53  
  
RESULT 9  
EL3B HUMAN STANDARD; PRT; 270 AA.  
ID EL3B HUMAN STANDARD; PRT; 270 AA.  
AC P08861; F11423;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Elastase IIIB precursor (EC 3.4.21.70) (Protease E).  
GN ELA3B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RX MEDLINE=88087253; PubMed=2826474;  
RA Tani T., Ohsumi J., Mita K., Takiguchi Y.;  
RT "Identification of a novel class of elastase isozyme, human  
RT pancreatic elastase III, by cDNA and genomic gene cloning.";  
RL J. Biol. Chem. 263:1231-1239(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RA Strausberg R.;

RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 4-270 FROM N.A.  
RC TISSUE=Pancreas;  
RX MEDLINE=88000545; PubMed=3477287;  
RA Shen W., Fletcher T.S., Largman C.;  
RT "Primary structure of human pancreatic protease E determined by  
RT sequence analysis of the cloned mRNA.";  
RL Biochemistry 26:3447-3452(1987).  
RN [4]  
RP SEQUENCE OF 31-50.  
RC TISSUE=Pancreas;  
RX MEDLINE=89325560; PubMed=2753124;  
RA Moulard M., Kerfelec B., Mallet B., Chapus C.;  
RT "Identification of a procarboxypeptidase A-truncated protease E  
RT binary complex in human pancreatic juice.";  
RL FEBS Lett. 250:166-170(1989).  
RN [5]  
RP SEQUENCE OF 94-164, AND CARBOHYDRATE-LINKAGE SITES ASN-114.  
RC TISSUE=Pancreas;  
RX MEDLINE=89289996; PubMed=2737288;  
RA Wendorf P., Geyer R., Sziegoleit A., Linder D.;  
RT "Localization and characterization of the glycosylation site of human  
RT pancreatic elastase 1.";  
RL FEBS Lett. 243:275-278(1989).  
CC -1- FUNCTION: EFFICIENT PROTEASE WITH ALANINE SPECIFICITY BUT ONLY  
CC LITTLE ELASTOLYTIC ACTIVITY.  
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Ala-Xaa. Does not  
CC hydrolyse elastin.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ELASTASE SUBFAMILY.  
CC -1- CAUTION: Was originally (Ref.5) thought to be elastase 1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M16630; AAA36482.1; -.  
DR EMBL; BC005216; AAH05216.1; -.  
DR EMBL; M18692; AAA58454.1; -.  
DR PIR; B29934; B29934.  
DR PIR; A27206; A27206.  
DR PIR; S04999; S04999.  
DR PIR; S04490; S04490.  
DR HSSP; P05805; IFON.  
DR MEROPS; S01.205; -.  
DR GlycoSuiteDB; P08861; -.  
DR SWISS-2DPAGE; P08861; HUMAN.  
DR Genew; HGNC:15945; ELA3B.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS50240; TRYPsin DOM; 1.  
DR PROSITE; PS00134; TRYPsin HIS; 1.  
DR PROSITE; PS00135; TRYPsin SER; 1.  
KW Hydrolase; Serine protease; Zymogen; Signal; Glycoprotein.  
FT SIGNAL 1 15  
FT PROPEP 16 28 ACTIVATION PEPTIDE (POTENTIAL).  
FT CHAIN 29 270  
FT ACT\_SITE 73 73 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 123 123 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 58 74 BY SIMILARITY.  
FT DISULFID 117 120 PROBABLE.  
FT DISULFID 157 223 BY SIMILARITY.  
FT DISULFID 188 204 BY SIMILARITY.  
FT DISULFID 213 244 BY SIMILARITY.  
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .).

```

FT CONFLICT 4 4 /FTID=CAR_000212.
FT CONFLICT 64 64 R -> G (IN REF. 3).
FT CONFLICT 79 79 A -> G (IN REF. 3).
FT CONFLICT 79 79 W -> R (IN REF. 1).
FT CONFLICT 129 131 MISSING (IN REF. 5).
FT CONFLICT 164 164 R -> P (IN REF. 3).
SQ SEQUENCE 270 AA; 23293 MW; B14BE0AAD369SAFE CRC64;

Query Match 83.0%; Score 93; DB 1; Length 270;
Best Local Similarity 84.2%; Pred. No. 2.4e-06;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPQVSLQ 19
Db 29 VNGEDAVPYSPWQVSLQ 47

RESULT 10
CLCR RAT
ID CLCR RAT STANDARD; PRT; 268 AA.
AC P5031; Q63148;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Caldecrin precursor (SC 3.4.21.2) (Chymotrypsin C) (serum calcium-
decreasing factor).
GN CTRC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP TISSUE=Pancreas;
RC MEDLINE=96107178; PubMed=8530454;
RA Tomomura A., Tomomura M., Fukushige T., Akiyama M., Kubota N.,
RA Kumaki K., Nishii Y., Noikura T., Saheki T.;
RT "Molecular cloning and expression of serum calcium-decreasing factor
(caldecrin).";
RL J. Biol. Chem. 270:30315-30321(1995).
RN [3]
RN SEQUENCE FROM N.A.
RP TISSUE=Pancreas;
RC MEDLINE=92165057; PubMed=1537555;
RA Kang J., Wiegand U., Mueller-Hill B.;
RT "Identification of cDNAs encoding two novel rat pancreatic serine
proteases.";
RL Gene 110:181-187(1992).
RN [3]
RN CHARACTERIZATION.
RP MEDLINE=98207038; PubMed=9538241;
RA Yoshino-Yasuda I., Kobayashi K., Akiyama M., Itoh H., Tomomura A.,
RA Saheki T.;
RT "Caldecrin is a novel-type serine protease expressed in pancreas, but
its homologue, elastase IV, is an artifact during cloning derived
from caldecrin gene.";
RL J. Biochem. 123:546-554(1998).
CC -1- FUNCTION: HAS CHYMOTRYPSIN-TYPE PROTEASE ACTIVITY AND HYPOCALCEMIC
ACTIVITY.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Leu-|-Xaa, Tyr-|-Xaa,
Phe-|-Xaa, Met-|-Xaa, Trp-|-Xaa, Gln-|-Xaa, Asn-|-Xaa.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.
CC -1- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE ELASTASE IV.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----

```

```

DR EMBL; S80379; AAB35830.1; -
DR EMBL; X59014; CAA41753.1; -
DR HSSP; P00766; 1CHG.
DR MEROPS; S01.157; -.
DR InterPro; IPR001314; Chymotrypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 16
FT PROPEP 17 29
FT CHAIN 30 268
FT ACT_SITE 74 74
FT ACT_SITE 123 123
FT ACT_SITE 216 216
FT DISULFID 17 141
FT DISULFID 59 75
FT DISULFID 155 222
FT DISULFID 186 202
FT DISULFID 212 243
FT CARBOHYD 25 25
FT CARBOHYD 90 90
FT CONFLICT 42 42
FT CONFLICT 96 120
FT SEQUENCE 268 AA; 29374 MW; 33B67AF34D0F8583 CRC64;
Query Match 80.4%; Score 90; DB 1; Length 268;
Best Local Similarity 78.9%; Pred. No. 6.4e-06;
Matches 15; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPQVSLQ 19
Db 30 VNGEDAVPYSPWQVSLQ 48

RESULT 11
EL3A HUMAN
ID EL3A HUMAN STANDARD; PRT; 270 AA.
AC P09053; Q9BRW4;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Elastase IIIA precursor (EC 3.4.21.70) (Protease E).
GN ELA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Pancreas;
RC MEDLINE=88087253; PubMed=2826474;
RA Tani T., Ohsumi J., Mita K., Takiguchi Y.;
RT "Identification of a novel class of elastase isozyme, human
pancreatic elastase III, by cDNA and genomic gene cloning.";
RL J. Biol. Chem. 263:1231-1239(1988).
RN [2]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP TISSUE=Pancreas;
RC MEDLINE=89034017; PubMed=2460440;
RA Shirasu Y., Takemura K., Yoshida H., Sato Y., Iijima H.,
RA Shimada Y., Mikayama T., Ozawa T., Ikeda N., Ishida A., Tanai Y.,
RA Matsuki S., Tanaka J., Ikenaga H., Ogawa M.;
RT "Molecular cloning of complementary DNA encoding one of the human
pancreatic protease E isozymes.";
RL J. Biochem. 104:259-264(1988).
RN [3]
RN SEQUENCE FROM N.A.

```

```

RC TISSUE=Prostate;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: EFFICIENT PROTEASE WITH ALAMINE SPECIFICITY BUT ONLY
CC LITTLE ELASTOLYTIC ACTIVITY.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Ala-|-Xaa. Does not
CC hydrolyse elastin.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M18700; AAA66350.1; -.
DR EMBL; M18693; AAA66350.1; JOINED.
DR EMBL; M18694; AAA66350.1; JOINED.
DR EMBL; M18695; AAA66350.1; JOINED.
DR EMBL; M18696; AAA66350.1; JOINED.
DR EMBL; M18697; AAA66350.1; JOINED.
DR EMBL; M18698; AAA66350.1; JOINED.
DR EMBL; M18699; AAA66350.1; JOINED.
DR EMBL; D00306; BAA00212.1; -.
DR EMBL; BC005918; AAH05918.1; -.
DR PIR; A29934; A29934.
DR HSP; P05805; IPON.
DR MEROPS; S01.154; -.
DR Genew; HGNC:15944; ELA3A.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal; Glycoprotein.
FT SIGNAL 1 15
FT PROPEP 16 28
FT CHAIN 29 270
FT ACT_SITE 73 73
FT ACT_SITE 123 123
FT ACT_SITE 217 217
FT DISULFID 58 74
FT DISULFID 117 120
FT DISULFID 157 223
FT DISULFID 188 204
FT DISULFID 213 244
FT CARBOHYD 114 114
FT CONFLICT 63 63
FT CONFLICT 106 106
FT CONFLICT 174 174
SQ SEQUENCE 270 AA; 576DDB255A4A118C CRC64;

Query Match 78.68; Score 88; DB 1; Length 270;
Best Local Similarity 78.94; Pred. No. 1.3e-05;
Matches 15; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 IVNGEEAVPGSWPWQVSLQ 19
Db 29 VVHGEDAVPYSPWPQVSLQ 47
:|||||
:|||||

RESULT 12
EL2A HUMAN
ID EL2A_HUMAN STANDARD; PRT; 269 AA.
AC P08217; O14243;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)

```

```

DE Elastase 2A precursor (EC 3.4.21.71).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87217962; PubMed=3646943;
RA Kawashima I., Tani T., Shimoda K., Takiguchi Y.;
RT "Characterization of pancreatic elastase II cDNAs: two elastase II
RL mRNAs are expressed in human pancreas.";
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98107669; PubMed=3427074;
RA Fletcher T.S., Shen W.F., Largman C.;
RT "Primary structure of human pancreatic elastase 2 determined by
RL sequence analysis of the cloned mRNA.";
RN [3]
RP SEQUENCE FROM N.A.
RX TISSUE=Pancreas;
RA Shirasu Y., Yoshida H., Matsuki S., Takemura K., Ikeda N., Shimada Y.,
RA Ozawa T., Mikayama T., Iijima H., Ishida A., Sato Y., Tamai Y.,
RA Tanaka J., Ikenaga H.;
RT "Molecular cloning and expression in Escherichia coli of a cDNA
RL encoding human pancreatic elastase 2.";
RN [4]
RP SEQUENCE FROM N.A.
RA Thomas D.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ACTS UPON ELASTIN.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Leu-|-Xaa, Met-|-Xaa
CC and Phe-|-Xaa. Hydrolyzes elastin.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16631; AAA52374.1; -.
DR EMBL; M16652; AAA52380.1; -.
DR EMBL; D00236; BAA00165.1; -.
DR EMBL; AL512883; CAC42421.1; -.
DR EMBL; BC007031; AAH07031.1; -.
DR PIR; A27432; A27432.
DR PIR; B26823; B26823.
DR HSP; P00772; IELG.
DR MEROPS; S01.155; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 16
FT PROPEP 17 28
FT ACTIVATION PEPTIDE.

```

```

FT CHAIN 29 269 ELASTASE 2A.
FT DISULFID 58 74 BY SIMILARITY.
FT DISULFID 155 222 BY SIMILARITY.
FT DISULFID 186 202 BY SIMILARITY.
FT DISULFID 212 243 BY SIMILARITY.
FT ACT SITE 73 73 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 121 121 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 216 216 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CONFLICT 202 202 C -> V (IN REF. 3).
SQ SEQUENCE 269 AA; 28888 MW; A2E05143EFF4987C CRC64;

Query Match 76.8%; Score 86; DB 1; Length 269;
Best Local Similarity 78.9%; Pred. No. 2.4e-05;
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPMQVSLQ 19
Db 29 VVGGEARPNPNSWPMQVSLQ 47

RESULT 13
ID_CAC3_BOVIN STANDARD; PRT; 253 AA.
AC P05805;
DT 01-NOV-1988 (Rel. 09, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proproteinase E precursor (Procarboxypeptidase A complex component
DE III) (Procarboxypeptidase A-S6 subunit III) (PROCPA-S6 III).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE OF 1-25.
RX MEDLINE=91099520; PubMed=2269366;
RA Pascual R., Vendrell J., Aviles F.X., Bonicel J., Wicker C.,
RA Puigserver A.;
RT Autolysis of proproteinase E in bovine procarboxypeptidase A ternary
RT complex gives rise to subunit III.";
RL FEBS Lett. 277:37-41(1990).
RN [2]
RP SEQUENCE OF 14-253, AND DISULFIDE BONDS.
RX MEDLINE=86220198; PubMed=3519215;
RA Venot N., Sciaky M., Puigserver A., Desmuelle P., Laurent G.;
RT "Amino acid sequence and disulfide bridges of subunit III, a
RT defective endopeptidase present in the bovine pancreatic 6 S
RT procarboxypeptidase A complex.";
RL Eur. J. Biochem. 157:91-99(1986).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RX MEDLINE=94222022; PubMed=8168476;
RA Pignol D., Gaboriaud C., Michon T., Kerfelec B., Chapus C.,
RA Fontecilla-Camps J.C.;
RT "Crystal structure of bovine procarboxypeptidase A-S6 subunit III, a
RT highly structured truncated zymogen E.";
RL EMBO J. 13:1763-1771(1994).
CC -1- FUNCTION: DEFECTIVE ELASTASE-LIKE SERINE PROTEASE. DOES NOT SEEM
CC TO HAVE A PROTEASE ACTIVITY. ITS LIKELY FUNCTION IS TO PROTECT
CC PROCARBOXYPEPTIDASE A AGAINST DENATURATION IN THE ACIDIC
CC ENVIRONMENT OF THE RUMINANT DUODENUM.
CC -1- SUBUNIT: HETEROTRIMER OF SUBUNIT III; CARBOXYPEPTIDASE A AND
CC CHYMOTRYPSINOGEN C.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR PIR; A25065; CPBOA3.
DR PDB; 1FON; 14-OCT-96.
DR MEROPS; S01.983; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser. protease_Try.
DR Pfam; PF00069; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYD_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Serine protease homolog; Pancreas; Digestion; 3D-structure.
FT PROPEP 1 11 ACTIVATION PEPTIDE.
FT CHAIN 12 253 PROTEINASE E.
FT DISULFID 41 57
FT DISULFID 100 103
FT DISULFID 140 206
FT DISULFID 171 187
FT DISULFID 196 227
SQ SEQUENCE 253 AA; 27337 MW; 24663724D8AE409C CRC64;

Query Match 75.9%; Score 85; DB 1; Length 253;
Best Local Similarity 78.9%; Pred. No. 3.2e-05;
Matches 15; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPMQVSLQ 19
Db 12 VVGGEAVPYWSWQVSLQ 30

RESULT 14
ID_EL2_RAT STANDARD; PRT; 271 AA.
AC P00774;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Elastase 2 precursor (EC 3.4.21.71).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82182967; PubMed=6918221;
RA McDonald R.J., Swift G.H., Quinto C., Swain W., Pictet R.L.,
RA Nikovits W., Rutter W.J.;
RT "Primary structure of two distinct rat pancreatic preproelastases
RT determined by sequence analysis of the complete cloned messenger
RT ribonucleic acid sequences.";
RL Biochemistry 21:1453-1463(1982).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85054882; PubMed=6094548;
RA Swift G.H., Craik C.S., Stary S.J., Quinto C., Lahaie R.G.,
RA Rutter W.J., MacDonald R.J.;
RT "Structure of the two related elastase genes expressed in the rat
RT pancreas.";
RL J. Biol. Chem. 259:14271-14278(1984).
CC -1- FUNCTION: ACTS UPON ELASTIN.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Leu-Xaa, Met-Xaa
CC and Phe-Xaa. Hydrolyzes elastin.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL curation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; V01233; CAA24543.1; -.
DR EMBL; L00124; AAA98780.1; -.
DR EMBL; L00118; AAA98780.1; JOINED.
DR EMBL; L00119; AAA98780.1; JOINED.

```

```

DR EMBL; L00120; AAA98780.1; JOINED.
DR EMBL; L00121; AAA98780.1; JOINED.
DR EMBL; L00122; AAA98780.1; JOINED.
DR EMBL; L00123; AAA98780.1; JOINED.
DR PIR; A00961; ELRT2.
DR HSP; P00772; 1ELG.
DR MEROPS; S01.155; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 16
FT CHAIN 17 30 ACTIVATION PEPTIDE.
FT CHAIN 31 271 ELASTASE 2.
FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 123 123 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 218 218 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 60 76 BY SIMILARITY.
FT DISULFID 157 224 BY SIMILARITY.
FT DISULFID 188 204 BY SIMILARITY.
FT DISULFID 214 245 BY SIMILARITY.
SQ SEQUENCE 271 AA; 28885 MW; 125C783B857B71E3 CRC64;

Query Match 75.0%; Score 84; DB 1; Length 271;
Best Local Similarity 73.7%; Pred. No. 4.7e-05;
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVNGEEAVPGSPWQVSLQ 19
   :|||:|||||
Db 31 VVGQERSPSPWQVSLQ 49

RESULT 15
TMS2 HUMAN
ID TMS2_HUMAN STANDARD; PRT; 492 AA.
AC O1533; Q9BX1;
DC 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transmembrane protease, serine 2 precursor (EC 3.4.21.-).
GN TMPRSS2 OR PRSS10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=97468144; PubMed=9325052;
RA Paoloni-Giacobino A., Chen H., Feitsch M.C., Rossier C.,
RA Antonarakis S.E.;
RT "Cloning of the TMPRSS2 gene, which encodes a novel serine protease
RT with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3.";
RL Genomics 44:309-320(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21309069; PubMed=11414763;
RA Teng D.H., Chen Y., Lian L., Ha P.C., Tavtigian S.V., Wong A.K.;
RT "Mutation analyses of 268 candidate genes in human tumor cell lines.";
RL Genomics 74:352-364(2001).
RN [3]
RP SEQUENCE FROM N.A. AND MUTAGENESIS.
RX MEDLINE=21139112; PubMed=11245484;
RA Afar D.E.H., Vivanco I., Hubert R.S., Kuo J., Chen E., Saffran D.C.,
RA Raitano A.B., Jakobovits A.;
RT "Catalytic cleavage of the androgen-regulated TMPRSS2 protease results
RT in its secretion by prostate and prostate cancer epithelia.";
RL Cancer Res. 61:1686-1692(2001).
RN [4]

TISSUE SPECIFICITY.
RP MEDLINE=21104370; PubMed=11169526;
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Vihko P.T.;
RT "Expression of transmembrane serine protease TMPRSS2 in mouse and
RT human tissues.";
RL J. Pathol. 193:134-140(2001).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY
CC CLEAVAGE AND SECRETED.
CC -1- TISSUE SPECIFICITY: EXPRESSED STRONGLY IN SMALL INTESTINE. ALSO
CC EXPRESSED IN PROSTATE, COLON, STOMACH, AND SALIVARY GLAND.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U75329; AAC51784.1; -.
DR EMBL; AF123453; AAD37117.1; -.
DR EMBL; AF270487; AAK29280.1; -.
DR HSP; P00763; 1DPO.
DR MEROPS; S01.247; -.
DR Genew; HGNC:11876; TMPRSS2.
DR MIM; 602060; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001272; LDL_recept_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Srcr_receptoi.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00058; LDLRA_2; 1.
DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS00287; SRCR_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
KW Polymorphism.
FT CHAIN 1 255 TRANSMEMBRANE PROTEASE, SERINE 2, NON-
FT CATALYTIC CHAIN.
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT CYTOPLASMIC (POTENTIAL).
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT EXTRACELLULAR (POTENTIAL).
FT LDL-RECEPTOR CLASS A.
FT SRCR.
FT SERINE PROTEASE.
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 255 256 CLEAVAGE (POTENTIAL).
FT SITE 113 126 BY SIMILARITY.
FT DISULFID 120 139 BY SIMILARITY.
FT DISULFID 133 148 BY SIMILARITY.
FT DISULFID 172 231 INTERCHAIN (BY SIMILARITY).
FT DISULFID 185 241 BY SIMILARITY.
FT DISULFID 244 365 BY SIMILARITY.
FT DISULFID 281 297 BY SIMILARITY.
FT DISULFID 410 426 BY SIMILARITY.
FT DISULFID 437 465 BY SIMILARITY.
FT CARBOHYD 213 249 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. .) (POTENTIAL).

```



FT VARIANT 449 449 K -> N (IN DBSNP:1056602).  
FT MUTAGEN 255 255 /FTid=VAR\_011692.  
FT MUTAGEN 441 441 R->Q: LOSS OF CLEAVAGE.  
FT CONFLICT 160 160 S->A: LOSS OF ACTIVITY.  
FT CONFLICT 242 242 M -> V (IN REF. 3).  
FT CONFLICT 329 329 I -> L (IN REF. 1).  
FT CONFLICT 489 491 E -> Q (IN REF. 1).  
SQ SEQUENCE 492 AA; 53891 MW; CAB44FD174A9076B CRC64;

Query Match 75.0%; Score 84; DB 1; Length 492;  
Best Local Similarity 77.8%; Pred. No. 8.4e-05;  
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVNGEEAVPGSWPQVSL 18  
Db 256 IVGESALFGAWPQVSL 273

Search completed: February 12, 2003, 10:23:18  
Job time : 3.68657 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:16:06 ; Search time 14.1493 Seconds  
(without alignments)  
291.248 Million cell updates/sec

Title: US-10-036-371-6

Perfect score: 112  
Sequence: 1 IYNGEEAVPGSWPQVSLQD 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organella.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_protist.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	108	96.4	263	11 Q9DC86	Q9dc86 mus musculus
2	108	96.4	263	11 Q9DX88	Q9dx88 mus musculus
3	108	96.4	263	11 Q9CR35	Q9cr35 mus musculus
4	107	95.5	264	11 Q9D960	Q9d960 mus musculus
5	107	95.5	264	11 Q9D7F8	Q9d7f8 mus musculus
6	107	95.5	264	11 Q9EQ28	Q9eq28 rattus norv
7	107	95.5	264	11 Q9ER05	Q9er05 mus musculus
8	101	90.2	260	13 Q9W703	Q9w7q3 paralichthy
9	98	87.5	263	13 Q9W704	Q9w7q4 paralichthy
10	98	87.5	263	13 Q9PW06	Q9pwq6 gadus morhu
11	97	86.6	269	11 Q9D7F9	Q9d7f9 mus musculus
12	97	86.6	269	11 Q9CQ52	Q9cq52 mus musculus
13	93	83.0	257	6 Q19023	Q19023 macaca muli
14	93	83.0	257	4 Q96QL8	Q96ql8 homo sapien
15	88	78.6	270	4 Q96KL7	Q96kl7 bos taurus
16	85	75.9	269	6 Q95KW7	

17	84	75.0	267	5 Q9BK47	Q9bk47 luidia foli
18	84	75.0	492	4 Q9ET73	Q9et73 homo sapien
19	81	72.3	20	13 Q9PRR4	Q9pr4 scyllorhinu
20	81	72.3	269	4 Q96QV5	Q96qv5 homo sapien
21	81	72.3	270	13 Q91039	Q91039 gadus morhu
22	80	71.4	20	13 Q9PRR3	Q9pr3 scyllorhinu
23	80	71.4	461	5 Q8T4N2	Q8t4n2 thipicephal
24	79	70.5	298	5 Q8T4N4	Q8t4n4 thipicephal
25	79	70.5	417	11 Q8VHK8	Q8vhk8 mus musculu
26	79	70.5	417	11 Q8VDV1	Q8vdv1 mus musculu
27	78	69.6	249	13 Q9W7Q1	Q9w7q1 paralichthy
28	78	69.6	266	13 Q9W7Q0	Q9w7q0 paralichthy
29	78	69.6	268	13 Q9W7Q2	Q9w7q2 paralichthy
30	77	68.8	1524	13 Q91674	Q91674 xenopus lae
31	76	67.9	421	11 Q60491	Q60491 cavia porce
32	76	67.9	474	5 Q8T4N3	Q8t4n3 thipicephal
33	75	67.0	377	6 P79343	P79343 bos taurus
34	75	67.0	405	4 Q96886	Q96886 homo sapien
35	74	66.1	274	11 Q924N9	Q924n9 mus musculu
36	74	66.1	310	11 Q90Y29	Q9gy29 mus musculu
37	74	66.1	310	11 Q91XC4	Q91xc4 mus musculu
38	74	66.1	490	11 Q920K3	Q920k3 rattus norv
39	73	65.2	277	5 Q96899	Q96899 scolopendra
40	73	65.2	329	13 Q42272	Q42272 xenopus lae
41	73	65.2	415	6 Q29015	Q29015 sus sp. pre
42	72	64.3	279	11 Q9QZ74	Q9qz74 rattus norv
43	72	64.3	329	6 Q9GL10	Q9gl10 ovis aries
44	72	64.3	417	11 Q8VHJ4	Q8vhj4 rattus norv
45	71	63.4	264	13 Q8QGF6	Q8qgf6 xenopus lae

## ALIGNMENTS

### RESULT 1

ID	Q9DC86	PRELIMINARY;	PRT;	263	AA.
AC	Q9DC86;				
DT	01-JUN-2001 (TRENBLrel. 17, Created)				
DT	01-JUN-2001 (TRENBLrel. 17, Last sequence update)				
DE	01-JUN-2002 (TRENBLrel. 21, Last annotation update)				
DE	220000809Rik protein.				
GN	220000809Rik.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=SPLLEN;				
RC	MEDLINE=21085660; PubMed=11217851;				
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,				
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,				
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,				
RA	Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,				
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Kasavant T.,				
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,				
RA	Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,				
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,				
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,				
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,				
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,				
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,				
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,				
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,				
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,				
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,				
RA	Yanishaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,				
RA	Hayashizaki Y.;				
RT	"Functional annotation of a full-length mouse cDNA collection.";				
PL	Nature 409:685-690 (2001).				
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE				
CC	TRYPsin FAMILY.				

```

DR EMBL; AK003060; BAB22539.1; -.
DR HSP; P00766; IGCT.
DR MEROPS; S01.152; -.
DR MGD; MGI:1913723; 2200008D09Rik
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPsin.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 263 AA; 27821 MW; 2620A27AFBA5D04D CRC64;

Query Match          96.4%; Score 108; DB 11; Length 263;
Best Local Similarity 90.0%; Pred. No. 2.5e-08;
Matches 18; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPQVSLQD 20
   |||||:|||||
Db 34 IVNGDAIPGSPWQVSLQD 53

RESULT 2
Q9D8X8
ID Q9D8X8 PRELIMINARY; PRT; 263 AA.
AC Q9D8X8
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 2200008D09Rik protein.
GN 2200008D09Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS7BL/6J; TISSUE=STOMACH, SPLEEN, AND PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Carninci P., de Bonaldo M.F.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayaishizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001)
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AK007566; BAB25112.1; -.
DR HSP; P00766; IGCT.
DR MEROPS; S01.152; -.
DR MGD; MGI:1913723; 2200008D09Rik.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPsin.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 263 AA; 27822 MW; 28C4487AF1A26827 CRC64;

DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 263 AA; 27898 MW; C0638FB8F905A92F CRC64;

Query Match          96.4%; Score 108; DB 11; Length 263;
Best Local Similarity 90.0%; Pred. No. 2.5e-08;
Matches 18; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVNGEAVPGSWPQVSLQD 20
   |||||:|||||
Db 34 IVNGDAIPGSPWQVSLQD 53

RESULT 3
Q9CR35
ID Q9CR35 PRELIMINARY; PRT; 263 AA.
AC Q9CR35
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 2200008D09Rik protein.
GN 2200008D09Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS7BL/6J; TISSUE=STOMACH, SPLEEN, AND PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Carninci P., de Bonaldo M.F.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayaishizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001)
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AK008927; BAB25971.1; -.
DR EMBL; AK003079; BAB22553.1; -.
DR EMBL; AK007765; BAB25241.1; -.
DR EMBL; AK007815; BAB25280.1; -.
DR EMBL; AK008729; BAB25861.1; -.
DR EMBL; AK008888; BAB25954.1; -.
DR HSP; P00766; IGCT.
DR MEROPS; S01.152; -.
DR MGD; MGI:1913723; 2200008D09Rik.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPsin.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 263 AA; 27822 MW; 28C4487AF1A26827 CRC64;

```

```

Query Match          96.4%; Score 108; DB 11; Length 263;
Best Local Similarity 90.0%; Pred. No. 2.5e-08;
Matches 18; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IVNGEAVPGSWPQVSLQD 20
    |||||:|||||:|||||
Db 34 IVNGDAIPGSPWQVSLQD 53

RESULT 4
Q9D960
ID Q9D960 PRELIMINARY; PRT; 264 AA.
AC Q9D960;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2002 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 1810004D15RIK protein.
GN CTRL OR 1810004D15RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Boffelli D., Bojunga N., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberto N.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszewski B., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AK009019; BAB26029.1; -.
DR HSP; P00766; 4CHA.
DR MEROPS; S01.256; -.
DR MGD; MGI:88558; Ctrl.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 264 AA; 28151 MW; 1D979719E07C16DE CRC64;

Query Match          95.5%; Score 107; DB 11; Length 264;
Best Local Similarity 95.0%; Pred. No. 3.6e-08;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IVNGEAVPGSWPQVSLQD 20
    |||||:|||||:|||||
Db 34 IVNGEAVPGSWPQVSLQD 53

RESULT 5
Q9EQ28
ID Q9EQ28 PRELIMINARY; PRT; 264 AA.
AC Q9EQ28;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2002 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Chymopasin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

```

```

Q9D7P8
ID Q9D7P8 PRELIMINARY; PRT; 264 AA.
AC Q9D7P8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 1810004D15RIK protein.
GN CTRL OR 1810004D15RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Boffelli D., Bojunga N., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberto N.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszewski B., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AK009019; BAB26029.1; -.
DR HSP; P00766; 4CHA.
DR MEROPS; S01.256; -.
DR MGD; MGI:88558; Ctrl.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 264 AA; 28167 MW; 1D979469A07056C2 CRC64;

Query Match          95.5%; Score 107; DB 11; Length 264;
Best Local Similarity 95.0%; Pred. No. 3.6e-08;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IVNGEAVPGSWPQVSLQD 20
    |||||:|||||:|||||
Db 34 IVNGEAVPGSWPQVSLQD 53

RESULT 6
Q9EQ28
ID Q9EQ28 PRELIMINARY; PRT; 264 AA.
AC Q9EQ28;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2002 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Chymopasin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

```

OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=RAT PANCREAS;  
 RA Sogame Y., Mitsui S., Kataoka K., Kashima K., Kato M., Sakagami J.,  
 RA Yamaguchi N.;  
 RT "Molecular cloning of rat chymopasin.";   
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 DR EMBL; AB020757; BAB20287.1; -.  
 DR HSP; P00766; 4CHA.  
 DR MEROPS; S01.256; -.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Serine protease.  
 SQ SEQUENCE 264 AA; 28116 MW; F9ED5D210FD3500E CRC64;

Query Match 95.5%; Score 107; DB 11; Length 264;  
 Best Local Similarity 95.0%; Pred. No. 3.6e-08;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IVNGEEAVPGSWPQVSLQD 20  
 Db 34 IVNGENAVPGSWPQVSLQD 53  
 ||||| ||||| ||||| ||||| |||||

RESULT 7  
 Q9ER05 PRELIMINARY; PRT; 264 AA.  
 AC Q9ER05;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Chymopasin (Chymotrypsin A CTRA-1).  
 GN CTRL OR CTRA1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Mitsui S., Yamaguchi N.;  
 RT "Molecular cloning of mouse chymopasin.";   
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 DR EMBL; AB016228; BAB20275.1; -.  
 DR EMBL; AF236365; AAL11034.1; -.  
 DR HSP; P00766; 4CHA.  
 DR MEROPS; S01.256; -.  
 DR MGD; MGI:88558; Ctrl.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Serine protease.  
 SQ SEQUENCE 264 AA; 28135 MW; 1D979709A07056C2 CRC64;

Query Match 95.5%; Score 107; DB 11; Length 264;  
 Best Local Similarity 95.0%; Pred. No. 3.6e-08;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 IVNGEEAVPGSWPQVSLQD 20  
 Db 34 IVNGENAVPGSWPQVSLQD 53  
 ||||| ||||| ||||| ||||| |||||

RESULT 8  
 Q9W7Q3 PRELIMINARY; PRT; 260 AA.  
 AC Q9W7Q3;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Chymotrypsinogen 2.  
 OS Paralichthys olivaceus (Flounder).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;  
 OC Pleuronectoidei; Paralichthyidae; Paralichthys.  
 OX NCBI\_TaxID=8255;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PANCREAS;  
 RA Suzuki T., Srivastava A.S., Kurokawa T.;  
 RT "Japanese flounder mRNA for chymotrypsinogen 2.";   
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 DR EMBL; AB029754; BAA82366.1; -.  
 DR HSP; P00766; 1CHG.  
 DR MEROPS; S01.152; -.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Serine protease.  
 SQ SEQUENCE 260 AA; 27793 MW; 9F583044E22F78C0 CRC64;

Query Match 90.2%; Score 101; DB 13; Length 260;  
 Best Local Similarity 90.0%; Pred. No. 2.8e-07;  
 Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 IVNGEEAVPGSWPQVSLQD 20  
 Db 31 IVNGEALPHSWPQVSLQD 50  
 ||||| ||||| ||||| ||||| |||||

RESULT 9  
 Q9W7Q4 PRELIMINARY; PRT; 261 AA.  
 AC Q9W7Q4;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Chymotrypsinogen 1.  
 OS Paralichthys olivaceus (Flounder).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;  
 OC Pleuronectoidei; Paralichthyidae; Paralichthys.  
 OX NCBI\_TaxID=8255;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PANCREAS;  
 RA Suzuki T., Srivastava A.S., Kurokawa T.;

RT "Japanese flounder mRNA for chymotrypsinogen 1.":  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.

DR EMBL: AB029753; BAA82365.1; --.

DR HSSP: P00763; IPOPO.

DR MEROPS: S01.256; --.

DR InterPro: IPR001314; Chymotrypsin.

DR InterPro: IPR001254; Ser. Protease\_Try.

DR Pfam: PF00089; trypsin; 1.

DR PRINTS: PR00722; CHYMOTRYPSIN.

DR SMART: SM00020; Tryp\_Spc; 1.

DR PROSITE: PS00240; TRYPSIN\_DOM; 1.

DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.

DR PROSITE: PS00135; TRYPSIN\_SER; 1.

KW Hydrolase; Serine protease.

SQ SEQUENCE 261 AA; 28184 MW; D7090A9D65395B7D CRC64;

Query Match 87.5%; Score 98; DB 13; Length 261;  
 Best Local Similarity 90.0%; Pred. No. 7.9e-07;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 INVGEAVPGSWPQVSLQ 20

Db 32 INVGETAVSGSWPQVSLQ 51

RESULT 10

ID Q9PW06 PRELIMINARY; PRT; 263 AA.

AC Q9PW06;

DT 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)

DE Chymotrypsin B precursor (EC 3.4.21.1).

GN CHYB.

OS Gadus morhua (Atlantic cod).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.

OC NCBI\_TaxID=8049;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=PYLORIC CAECA;

RX MEDLINE=20464334; PubMed=11011764;

RA Spilliaert R., Gudmundsdottir A.;

RT "Molecular Cloning of the Atlantic Cod Chymotrypsinogen B.:"

RL Microb. Comp. Genomics 5:41-50(2000).

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

CC TRYPSIN FAMILY.

DR EMBL: AJ242521; CAB43766.1; --.

DR HSSP: P00766; 1CHG.

DR MEROPS: S01.152; --.

DR InterPro: IPR001314; Chymotrypsin.

DR InterPro: IPR001254; Ser. Protease\_Try.

DR Pfam: PF00089; trypsin; 1.

DR PRINTS: PR00722; CHYMOTRYPSIN.

DR SMART: SM00020; Tryp\_Spc; 1.

DR PROSITE: PS00240; TRYPSIN\_DOM; 1.

DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.

DR PROSITE: PS00135; TRYPSIN\_SER; 1.

KW Hydrolase; Serine protease; Signal.

FT SIGNAL 1 16 POTENTIAL.

FT CHAIN 17 263 CHYMOTRYPSIN B.

SQ SEQUENCE 263 AA; 28175 MW; EF61B19A34EE57C CRC64;

Query Match 87.5%; Score 98; DB 13; Length 263;  
 Best Local Similarity 94.7%; Pred. No. 7.9e-07;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 INVGEAVPGSWPQVSLQ 19

Db 32 INVGEAVPHSWPQVSLQ 50

RESULT 11

Q9D7T9

ID Q9D7T9 PRELIMINARY; PRT; 269 AA.

AC Q9D7T9;

DT 01-JUN-2001 (Tremblrel. 17, Created)

DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE 2310074F01Rik protein.

GN ELA3B OR 2310074F01RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=STOMACH;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,

RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo V.F.,

RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,

RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.:"

RL Nature 409:685-690(2001).

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

CC TRYPSIN FAMILY.

DR EMBL: AK008858; BAB25932.1; --.

DR HSSP: P03805; IFON.

DR MEROPS: S01.154; --.

DR MGD: MGI:1915118; Ela3b.

DR InterPro: IPR001314; Chymotrypsin.

DR InterPro: IPR001254; Ser. Protease\_Try.

DR Pfam: PF00089; trypsin; 1.

DR PRINTS: PR00722; CHYMOTRYPSIN.

DR SMART: SM00020; Tryp\_Spc; 1.

DR PROSITE: PS00240; TRYPSIN\_DOM; 1.

DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.

DR PROSITE: PS00135; TRYPSIN\_SER; 1.

KW Hydrolase; Serine protease.

SQ SEQUENCE 269 AA; 28977 MW; 9F43F769DB2A7CF CRC64;

Query Match 86.6%; Score 97; DB 11; Length 269;

Best Local Similarity 89.5%; Pred. No. 1.1e-06;

Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 INVGEAVPGSWPQVSLQ 19

Db 28 VVNGEAVPHSWPQVSLQ 46

RESULT 12

Q9CQ52

ID Q9CQ52 PRELIMINARY; PRT; 269 AA.

AC Q9CQ52;

DT 01-JUN-2001 (Tremblrel. 17, Created)

DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE 2310074F01Rik protein.

GN ELA3B OR 2310074F01RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RX NCBI\_TaxID=10090;  
RN [1]\_SEQUENCE FROM N.A.  
RP STRAIN=C57BL/6J; TISSUE=TONGUE;  
RC MEDLINE=21085660; PubMed=1217851;  
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Akaiwa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.P.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,  
RA Suzuki H., Toyono-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtauki S.,  
RA Hayashizaki Y.;  
RA "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
DR EMBL; AK010149; BAB26734.1; -;  
DR EMBL; AK009129; BAB26092.1; -;  
DR HSSP; P05805; 1FON.  
DR MEROPS; S01.154; -;  
DR MGD; MGI:1915118; ELA3b.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser. protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease.  
SQ SEQUENCE 285 AA; 28904 MW; C543P76957B2A7CE CRC64;  
Query Match 86.6%; Score 97; DB 11; Length 269;  
Best Local Similarity 89.5%; Pred. No. 1.1e-06;  
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 IVNGEAVPGSWPQVSLQ 19  
Db 28 VVNGEAVPHSWPQVSLQ 46  
:|||||:|||||  
RESULT 13  
ID Q9PSP2 PRELIMINARY; PRT; 40 AA.  
AC Q9PSP2;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Chymotrypsin C, PI isoform (Fragments).  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]\_SEQUENCE.  
RP MEDLINE=95066525; PubMed=7976036;  
RX Tluscik F., Polanowski A., Guyonnet V., Long P.L., Travis J.;

RT "Affinity purification of chicken pancreas proteinases and their N-  
RT terminal amino-acid sequences.";  
RL Acta Biochim. Pol. 41:174-177(1994).  
DR HSSP; P00766; 1CHG.  
KW Hydrolase; Serine protease.  
FT NON\_TER 1  
FT NON\_CONS 16 17  
FT NON\_TER 40 40  
SQ SEQUENCE 40 AA; 4376 MW; 348CFFBFC8C8003D CRC64;  
Query Match 84.8%; Score 95; DB 13; Length 40;  
Best Local Similarity 89.5%; Pred. No. 3e-07;  
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 IVNGEAVPGSWPQVSLQ 19  
Db 17 IVHGEFVPGSWPQVSLQ 35  
:|||||:|||||  
RESULT 14  
ID O19023 PRELIMINARY; PRT; 257 AA.  
AC O19023;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Elastase (EC 3.4.21.36) (Fragment).  
DN ELAI.  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;  
OC Cercopitheciae; Macaca.  
OX NCBI\_TaxID=9544;  
RN [1]\_SEQUENCE FROM N.A.  
RP TISSUE=PANCREAS;  
RC Szigoleit A.;  
RA "Pancreatic elastase from rhesus monkey";  
RL Submitted (JUN-1997) to the EMBL/Genbank/DBSJ databases.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
DR EMBL; AJ000067; CAA03899.1; -;  
DR HSSP; P05805; 1FON.  
DR MEROPS; S01.154; -;  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser. protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease.  
FT NON\_TER 1  
FT NON\_TER 257 257  
SQ SEQUENCE 257 AA; 27687 MW; 4D443DB67233D8DC CRC64;  
Query Match 83.0%; Score 93; DB 6; Length 257;  
Best Local Similarity 84.2%; Pred. No. 4.3e-06;  
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 IVNGEAVPGSWPQVSLQ 19  
Db 16 VVNGEAVPVSWPQVSLQ 34  
:|||||:|||||  
RESULT 15  
ID Q96QL8 PRELIMINARY; PRT; 270 AA.  
AC Q96QL8;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)



DE Similar to elastase 3, pancreatic (protease E).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBT\_TaxID=9606;  
 RN [1]\_

RN SEQUENCE FROM N.A.  
 RC TISSUE=PANCREAS;  
 RA Strausberg R.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC008383; AAH08383.1; -.  
 DR MEROPS; S01.154; -.  
 DR InterPro: IPR001254; Ser\_Protease\_Try.  
 DR Pfam: PF00089; trypsin; 1.  
 DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; UNKNOWN\_1.  
 KW Hydrolase; Protease; Serine protease.  
 SQ SEQUENCE 270 AA; 29446 MW; 946DDDA694A102E CRC64;

Query Match 78.6%; Score 88; DB 4; Length 270;  
 Best Local Similarity 78.9%; Pred. No. 2.6e-05;  
 Matches 15; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IVNGEEAVPGSWPQVSLQ 19  
 Db 29 VVHGEDAVPYSPWPQVSLQ 47

Search completed: February 12, 2003, 10:27:26  
 Job time : 15.1493 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:18:40 ; Search time 4.29851 Seconds  
(without alignments)  
136.898 Million cell updates/sec

Title: US-10-036-371-6

Perfect score: 112

Sequence: 1 IVNGEEAVFGSWPQVSLQD 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 2942292 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

1: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep.\*

2: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep.\*

3: /cgn2\_6/prodata/1/iaa/6A\_COMB.pep.\*

4: /cgn2\_6/prodata/1/iaa/6B\_COMB.pep.\*

5: /cgn2\_6/prodata/1/iaa/PCTUS\_COMB.pep.\*

6: /cgn2\_6/prodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112	100.0	228	1	US-08-278-091-10
2	112	100.0	228	1	US-08-483-859-10
3	112	100.0	228	1	US-08-472-173-10
4	112	100.0	228	2	US-08-487-167-10
5	112	100.0	228	2	US-08-482-816-10
6	112	100.0	228	2	US-08-296-149-10
7	112	100.0	228	2	US-08-801-499-10
8	112	100.0	228	3	US-08-615-271-10
9	112	100.0	228	3	US-09-074-660-10
10	112	100.0	228	3	US-09-074-659-10
11	112	100.0	228	3	US-09-106-468-10
12	112	100.0	228	3	US-09-106-466A-10
13	112	100.0	228	4	US-09-106-467-10
14	109	97.3	20	2	US-08-385-540A-16
15	109	97.3	20	2	US-08-600-273A-16
16	109	97.3	20	3	US-08-486-820-16
17	109	97.3	20	4	US-09-220-731-16
18	109	97.3	229	2	US-08-557-146-13
19	109	97.3	229	2	US-09-154-344-13
20	109	97.3	230	4	US-08-944-483-62
21	109	97.3	231	2	US-09-027-337-6
22	109	97.3	231	4	US-09-644-600-6
23	93	83.0	31	5	PCT-US95-16826-1
24	93	83.0	32	5	PCT-US95-16826-2
25	93	83.0	242	4	US-08-944-483-57
26	88	78.6	242	4	US-08-944-483-58
27	86	76.8	241	4	US-08-944-483-59

28	84	75.0	283	3	US-08-807-151-1	Sequence 1, Appli
29	84	75.0	283	4	US-09-478-957-1	Sequence 1, Appli
30	84	75.0	492	4	US-09-342-749-2	Sequence 2, Appli
31	84	75.0	492	4	US-09-691-840-2	Sequence 2, Appli
32	81	72.3	241	4	US-08-944-483-60	Sequence 60, Appli
33	80	71.4	239	4	US-08-944-483-61	Sequence 61, Appli
34	80	71.4	268	1	US-08-270-584A-2	Sequence 2, Appli
35	80	71.4	268	1	US-08-568-031-2	Sequence 2, Appli
36	80	71.4	268	2	US-08-765-192-2	Sequence 2, Appli
37	80	71.4	268	2	US-08-966-319-2	Sequence 2, Appli
38	80	71.4	268	3	US-09-153-304-2	Sequence 2, Appli
39	80	71.4	268	3	US-09-199-793-2	Sequence 2, Appli
40	75	67.0	423	4	US-09-656-002-2	Sequence 2, Appli
41	75	67.0	435	4	US-09-008-271A-6	Sequence 6, Appli
42	73	65.2	20	1	US-08-508-448C-1	Sequence 1, Appli
43	73	65.2	232	1	US-08-508-448C-19	Sequence 19, Appli
44	73	65.2	246	4	US-09-370-818-60	Sequence 60, Appli
45	73	65.2	418	1	US-08-508-448C-25	Sequence 25, Appli

## ALIGNMENTS

RESULT 1  
US-08-278-091-10  
; Sequence 10, Application US/08278091  
; Patent No. 5506139  
; GENERAL INFORMATION:  
; APPLICANT: LOOMORE, Sheena M  
; APPLICANT: YANG, Yan-Ping  
; APPLICANT: CHONG, Pele  
; APPLICANT: COMEN, Raymond P.  
; APPLICANT: KLEIN, Michel H.  
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with  
; TITLE OF INVENTION: Reduced Protease Activity  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/278,091  
; FILING DATE: 21-JUL-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-371  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 228 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-278-091-10

Query Match 100.0%; Score 112; DB 1; Length 228;  
Best Local Similarity 100.0%; Pred. No 1.1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVNGEEAVFGSWPQVSLQD 20  
|||||

```
Db      1  IVNGEEAVPGSWPQVSLQD 20

RESULT 2
US-08-483-859-10
; Sequence 10, Application US/08483859
; Patent No. 5656436
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COHEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,859
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-493 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-472-173-10

Query Match      100.0%; Score 112; DB 1; Length 228;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1  IVNGEEAVPGSWPQVSLQD 20
| | | | | | | | | | | | | | | |
Db      1  IVNGEEAVPGSWPQVSLQD 20

RESULT 4
US-08-487-167-10
; Sequence 10, Application US/08487167
; Patent No. 5869302
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COHEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

Db      1  IVNGEEAVPGSWPQVSLQD 20

RESULT 3
US-08-472-173-10
; Sequence 10, Application US/08472173
; Patent No. 5665353
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COHEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
```

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,167
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-508 MIS:vg
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-487-167-10
;
; Query Match 100.0%; Score 112; DB 2; Length 228;
; Best Local Similarity 100.0%; Pred. No. 1.1e-09;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 IVNGEEAVPGSWPWQVSLQD 20
; DB 1 IVNGEEAVPGSWPWQVSLQD 20
;
; RESULT 5
; US-08-482-816-10
; Sequence 10, Application US/08482816
; Patent No. 5935573
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-494 MIS:vg
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-482-816-10
;
; Query Match 100.0%; Score 112; DB 2; Length 228;
; Best Local Similarity 100.0%; Pred. No. 1.1e-09;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 IVNGEEAVPGSWPWQVSLQD 20
; DB 1 IVNGEEAVPGSWPWQVSLQD 20
;
; RESULT 6
; US-08-296-149-10
; Sequence 10, Application US/08296149
; Patent No. 5939297
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 08/08/296,149
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-390
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-296-149-10
;
; Query Match 100.0%; Score 112; DB 2; Length 228;
; Best Local Similarity 100.0%; Pred. No. 1.1e-09;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 IVNGEEAVPGSWPWQVSLQD 20
; DB 1 IVNGEEAVPGSWPWQVSLQD 20
;
; RESULT 7

```

```
US-08-801-499-10
; Sequence 10, Application US/08801499
; Patent No. 5962430
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COHEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced Protease Act
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,499
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/482,816
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-671 MJS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-801-499-10

Query Match 100.0%; Score 112; DB 2; Length 228;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IVNGEEAVPGSWPWQVSLQD 20
Db 1 IVNGEEAVPGSWPWQVSLQD 20

RESULT 8
US-08-615-271-10
; Sequence 10, Application US/08615271
; Patent No. 5981503
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COHEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,660
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,167
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:

US-08-801-499-10
; Sequence 10, Application US/08801499
; Patent No. 5962430
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COHEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,499
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/482,816
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-671 MJS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-801-499-10

Query Match 100.0%; Score 112; DB 2; Length 228;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IVNGEEAVPGSWPWQVSLQD 20
Db 1 IVNGEEAVPGSWPWQVSLQD 20

RESULT 9
US-09-074-660-10
; Sequence 10, Application US/09074660
; Patent No. 6020183
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COHEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,660
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,167
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
```

```

: APPLICATION NUMBER: US 08/296,149
: FILING DATE: 26-AUG-1994
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/278,091
: FILING DATE: 21-JUL-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Stewart, Michael I.
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-731 MIS:jfb
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 228 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-09-074-660-10

Query Match 100.0% Score 112;
Best Local Similarity 100.0%; Pred No. 1,11
Matches 20; Conservative 0; Mismatches

QY 1 IVNGEEAVPGSPWQVSLQD 20
|||||
Db 1 IVNGEEAVPGSPWQVSLQD 20

RESULT 10
US-09-074-659-10
: Sequence 10, Application US/09074659
: Patent No. 6025342
: GENERAL INFORMATION:
: APPLICANT: LOOSMORE, Sheena M.
: APPLICANT: YANG, Yan-Ping
: APPLICANT: CHONG, Pele
: APPLICANT: KOENEN, Raymond P.
: APPLICANT: KLEIN, Michel H.
: TITLE OF INVENTION: Analog of Haemophilus F.
: TITLE OF INVENTION: Reduced Protease Active
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: 6th Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/074,659
: FILING DATE:
: CLASSIFICATION:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/487,167
: FILING DATE: 07-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/296,149
: FILING DATE: 26-AUG-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/278,091
: FILING DATE: 21-JUL-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Stewart, Michael I.
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-730 MIS:jfb
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155

```

```

; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-074-659-10

Query Match 100.0%; Score 112; DB 3; Length 228;
Best Local Similarity 100.0%; Pred.No.1.1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0

Qy 1 IVNGEEAVPGSWPQVSLQD 20
Db 1 IVNGEEAVPGSWPQVSLQD 20

RESULT 11
US-09-106-468-10
; Sequence 10, Application US/09106468
; Patent No. 6114125
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
; TITLE OF INVENTION: PROTEASE ACTIVITY
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106.468
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/615,271
; FILING DATE: 20-JUN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-825
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-106-468-10

Query Match 100.0%; Score 112; DB 3; Length 228;
Best Local Similarity 100.0%; Pred.No.1.1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0

Qy 1 IVNGEEAVPGSWPQVSLQD 20
Db 1 IVNGEEAVPGSWPQVSLQD 20

```

Wed Feb 12 11:59:42 2003

```
RESULT 12
US-09-106-466A-10
; Sequence 10, Application US/09106466A
; Patent No. 6147057
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
; TITLE OF INVENTION: PROTEASE ACTIVITY
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,466A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/615,271
; FILING DATE: 20-JUN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-826
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-106-466A-10
;
; Query Match 100.0%; Score 112; DB 4; Length 228;
; Best Local Similarity 100.0%; Pred. No. 1.1e-09;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
Qy 1 IVNGEEAVPGSWPQVSLQD 20
| | | | | | | | | | | | | | | | | | | | |
Db 1 IVNGEEAVPGSWPQVSLQD 20
| | | | | | | | | | | | | | | | | | | | |

RESULT 13
US-09-106-467-10
; Sequence 10, Application US/09106467
; Patent No. 6153580
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
; TITLE OF INVENTION: PROTEASE ACTIVITY
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,467
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/615,271
; FILING DATE: 20-JUN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-826
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-106-467-10
;
; Query Match 100.0%; Score 112; DB 4; Length 228;
; Best Local Similarity 100.0%; Pred. No. 1.1e-09;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
Qy 1 IVNGEEAVPGSWPQVSLQD 20
| | | | | | | | | | | | | | | | | | | | |
Db 1 IVNGEEAVPGSWPQVSLQD 20
| | | | | | | | | | | | | | | | | | | | |

RESULT 14
US-08-385-540A-16
; Sequence 16, Application US/08385540A
; Patent No. 5945102
; GENERAL INFORMATION:
; APPLICANT: de Faire, Johan
; TITLE OF INVENTION: Wound Care With Multifunctional
; TITLE OF INVENTION: Enzyme
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/385,540A
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/600,273
; FILING DATE: 08-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
```



REGISTRATION NUMBER: 29,135  
REFERENCE/DOCKET NUMBER: 314572-101A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 609-520-3214  
TELEFAX: 609-520-3259  
TELEX:  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-385-540A-16

Query Match 97.3%; Score 109; DB 2; Length 20;  
Best Local Similarity 95.0%; Pred. No. 1.9e-10;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVNGEEAVGSGMPWQVSLQD 20  
Db 1 IVNGEDAVGSGMPWQVSLQD 20

RESULT 15  
US-08-600-273A-16  
Sequence 16, Application US/08600273A  
Patent No 5958406  
GENERAL INFORMATION:  
APPLICANT: de Faire, Johan  
APPLICANT: Franklin, Richard L.  
APPLICANT: Kay, John  
TITLE OF INVENTION: Acne Treatment With Multifunctional  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert Price & Rhoads  
STREET: 997 Lenox Drive, Building 3, Suite 210  
CITY: Lawrenceville  
STATE: NJ  
COUNTRY: USA  
ZIP: 08543  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/600,273A  
FILING DATE: 08-FEB-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/486,820  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/385,540  
FILING DATE: 08-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Bloom, Allen  
REGISTRATION NUMBER: 29,135  
REFERENCE/DOCKET NUMBER: 314572-101C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 609-620-3214  
TELEFAX: 609-620-3259  
TELEX:  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-600-273A-16

Query Match 97.3%; Score 109; DB 2; Length 20;

Best Local Similarity 95.0%; Pred. No. 1.9e-10;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 IVNGEEAVGSGMPWQVSLQD 20  
Db 1 IVNGEDAVGSGMPWQVSLQD 20  
Search completed: February 12, 2003, 10:30:10  
Job time : 4.29851 secs

**THIS PAGE BLANK (uspto)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:22:36 ; Search time 3.04478 Seconds  
(without alignments)  
167.821 Million cell updates/sec

Title: US-10-036-371-6

Perfect score: 112

Sequence: 1 IVNGEEAVPGSWPQVSLQD 20

Scoring table: BL0SUM62

Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08 NEW PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT NEW PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06 NEW PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07 NEW PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09 NEW PUB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10 NEW PUB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60 NEW PUB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112	100.0	20	12 US-10-036-371-6	Sequence 6, Appli
2	109	97.3	146	10 US-09-925-297-643	Sequence 643, App
3	109	97.3	263	10 US-09-888-615-96	Sequence 96, Appl
4	97	86.6	192	10 US-09-925-297-529	Sequence 529, App
5	88	78.6	270	10 US-09-923-779-152	Sequence 152, App
6	86	76.8	269	10 US-09-925-297-576	Sequence 576, App
7	86	76.8	273	10 US-09-925-297-695	Sequence 695, App
8	85	75.9	20	12 US-10-036-371-4	Sequence 4, Appli
9	84	75.0	283	10 US-09-988-975A-1	Sequence 1, Appli
10	84	75.0	384	9 US-09-981-353-23	Sequence 23, Appli
11	84	75.0	393	9 US-10-012-896-934	Sequence 934, App
12	84	75.0	393	9 US-09-895-793-934	Sequence 934, App
13	84	75.0	393	9 US-09-895-814-934	Sequence 934, App
14	84	75.0	393	10 US-09-759-143-934	Sequence 934, App
15	84	75.0	393	10 US-09-780-669-934	Sequence 934, App
16	84	75.0	393	10 US-09-822-827-934	Sequence 934, App
17	84	75.0	492	9 US-10-012-896-895	Sequence 895, App
18	84	75.0	492	9 US-10-012-896-932	Sequence 932, App
19	84	75.0	492	9 US-09-895-793-895	Sequence 895, App

20	84	75.0	492	9 US-09-895-793-932	Sequence 932, App
21	84	75.0	492	9 US-09-895-814-895	Sequence 895, App
22	84	75.0	492	9 US-09-895-814-932	Sequence 932, App
23	84	75.0	492	10 US-09-759-143-895	Sequence 895, App
24	84	75.0	492	10 US-09-759-143-932	Sequence 932, App
25	84	75.0	492	10 US-09-780-669-895	Sequence 895, App
26	84	75.0	492	10 US-09-780-669-932	Sequence 932, App
27	84	75.0	492	10 US-09-822-827-895	Sequence 895, App
28	84	75.0	492	10 US-09-822-827-932	Sequence 932, App
29	84	75.0	492	10 US-09-822-827-932	Sequence 932, App
30	81	72.3	113	10 US-09-925-297-871	Sequence 871, App
31	76	67.9	288	10 US-09-888-615-115	Sequence 115, App
32	76	67.9	818	10 US-09-888-615-111	Sequence 111, App
33	75	67.0	406	10 US-09-851-588-6	Sequence 6, Appli
34	75	67.0	432	9 US-10-063-547-112	Sequence 112, App
35	75	67.0	432	9 US-10-174-590-330	Sequence 330, App
36	75	67.0	432	9 US-10-176-758-330	Sequence 330, App
37	75	67.0	432	9 US-10-063-616-112	Sequence 112, App
38	75	67.0	432	9 US-10-175-737-330	Sequence 330, App
39	75	67.0	432	9 US-10-063-502-112	Sequence 112, App
40	75	67.0	432	9 US-10-173-706-330	Sequence 330, App
41	75	67.0	432	9 US-10-175-738-330	Sequence 330, App
42	75	67.0	432	9 US-10-175-752-330	Sequence 330, App
43	75	67.0	432	9 US-10-176-482-330	Sequence 330, App
44	75	67.0	432	9 US-10-176-757-330	Sequence 330, App
45	75	67.0	432	9 US-10-176-913-330	Sequence 330, App

## ALIGNMENTS

RESULT 1  
US-10-036-371-6  
; Sequence 6, Application US/10036371  
; Patent No. US20020141987A1  
; GENERAL INFORMATION:  
; APPLICANT: BJARNARSON, JON B.  
; TITLE OF INVENTION: FISH SERINE PROTEINASES AND THEIR PHARMACEUTICAL AND  
; FILE REFERENCE: 81691/284960  
; CURRENT APPLICATION NUMBER: US/10/036,371  
; CURRENT FILING DATE: 2002-01-07  
; PRIOR APPLICATION NUMBER: 09/411,688  
; PRIOR FILING DATE: 1999-10-12  
; PRIOR APPLICATION NUMBER: 5086/99  
; PRIOR FILING DATE: 1999-06-18  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Bovine sp.  
US-10-036-371-6

Query Match 100.0%; Score 112; DB 12; Length 20;  
Best Local Similarity 100.0%; Pred. No. 7.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVNGEEAVPGSWPQVSLQD 20  
| | | | | | | | | | | | | | | | | | | | | |  
DB 1 IVNGEEAVPGSWPQVSLQD 20

RESULT 2  
US-09-925-297-643  
; Sequence 643, Application US/09925297  
; Patent No. US20020081659A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA105  
; CURRENT APPLICATION NUMBER: US/09/925,297  
; CURRENT FILING DATE: 2001-08-10



; CURRENT APPLICATION NUMBER: US/09/925,297  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05989  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 576  
; LENGTH: 269  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (8)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (167)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (213)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (220)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (234)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-297-576

Query Match 76.8%; Score 86; DB 10; Length 269;  
Best Local Similarity 78.9%; Pred. No. 4.3e-05;  
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 IVNGEEAVPGSWPQVSLQ 19  
:|||||  
Db 34 VVGGEARPNPWPQVSLQ 52

RESULT 7  
US-09-925-297-695  
; Sequence 695, Application US/09925297  
; Patent No. US20020081659A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA105  
; CURRENT APPLICATION NUMBER: US/09/925,297  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05989  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 695  
; LENGTH: 273  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (27)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (28)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (34)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-297-695

Query Match 76.8%; Score 86; DB 10; Length 273;  
Best Local Similarity 78.9%; Pred. No. 4.4e-05;

Matches 15; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 IVNGEEAVPGSWPQVSLQ 19  
:|||||  
Db 32 VVGGEDAVPYSWPQVSLQ 50

RESULT 8  
US-10-036-371-4  
; Sequence 4, Application US/10036371  
; Patent No. US20020141987A1  
; GENERAL INFORMATION:  
; APPLICANT: BJARNARSON, JON B.  
; TITLE OF INVENTION: FISH SERINE PROTEINASES AND THEIR PHARMACEUTICAL AND  
; FILE REFERENCE: 81691/284960  
; CURRENT APPLICATION NUMBER: US/10/036,371  
; CURRENT FILING DATE: 2002-01-07  
; PRIOR APPLICATION NUMBER: 09/411,688  
; PRIOR FILING DATE: 1999-10-12  
; PRIOR APPLICATION NUMBER: 5086/99  
; PRIOR FILING DATE: 1999-06-18  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Gadus sp.  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (11)  
; OTHER INFORMATION: S or T  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (13)  
; OTHER INFORMATION: S, P or Y  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (20)  
; OTHER INFORMATION: D or Q  
US-10-036-371-4

Query Match 75.9%; Score 85; DB 12; Length 20;  
Best Local Similarity 84.2%; Pred. No. 4.6e-06;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 IVNGEEAVPGSWPQVSLQ 19  
:|||||  
Db 1 IVNGEEAVPHXWQVSLQ 19

RESULT 9  
US-09-988-975A-1  
; Sequence 1, Application US/09988975A  
; Patent No. US20020119531A1  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Lal, Preeti G.  
; TITLE OF INVENTION: PROSTATE-ASSOCIATED PROTEASE ANTIBODY  
; FILE REFERENCE: PF-0227-2 CIP  
; CURRENT APPLICATION NUMBER: US/09/988,975A  
; CURRENT FILING DATE: 2001-11-19  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PERL Program  
; SEQ ID NO 1  
; LENGTH: 283  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc:feature  
; OTHER INFORMATION: Incyte ID No. US20020119531A1 556016  
; NAME/KEY: unsure



```
; Sequence 934, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 934
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-814-934

Query Match          75.0%; Score 84; DB 9; Length 393;
Best Local Similarity 77.8%; Pred. No. 0.00012;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVNGEEAVPGSWPQVSL 18
Db 157 IVGESALPGAMPQVSL 174

RESULT 14
US-09-759-143-934
; Sequence 934, Application US/09759143
; Patent No. US200202249A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; APPLICANT: Houghton, Raymond L.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C24
; CURRENT APPLICATION NUMBER: US/09/780,669
; CURRENT FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 943
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 934
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-780-669-934

Query Match          75.0%; Score 84; DB 10; Length 393;
Best Local Similarity 77.8%; Pred. No. 0.00012;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IVNGEEAVPGSWPQVSL 18
Db 157 IVGESALPGAMPQVSL 174

Search completed: February 12, 2003, 10:31:06
Job time : 3.04478 secs
```

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:03:55 ; Search time 2.68657 Seconds  
(without alignments)  
198.395 Million cell updates/sec

Title: US-10-036-371-7

Perfect score: 21

Sequence: 1 AAPF 4

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A. Geneseq\_101002.\*

```
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	4	5 AAP40459	N-terminal polyhyd
2	21	100.0	4	11 AAR04012	Peptidase substrat
3	21	100.0	4	12 AAR10887	Peptide component
4	21	100.0	4	12 AAR11223	Ketone analogue pr
5	21	100.0	4	13 AAR29964	Cathepsin G inhibi
6	21	100.0	4	14 AAR34258	Chromogenic peptid
7	21	100.0	4	14 AAR38410	Cathepsin G inhibi
8	21	100.0	4	14 AAR44109	Aminonaphthalene p
9	21	100.0	4	15 AAR52024	Bacillus alkali pr
10	21	100.0	4	15 AAR46223	Serine protease in

11	21	100.0	4	15 AAR53781	Sequence of intern
12	21	100.0	4	16 AAR72920	Substrate for pept
13	21	100.0	4	16 AAR72876	Substrate for pept
14	21	100.0	4	16 AAR77315	Porphyromonas ging
15	21	100.0	4	16 AAR77196	Cell proliferation
16	21	100.0	4	17 AAR85709	Degradable peptide
17	21	100.0	4	18 AAR52606	Serine protease-in
18	21	100.0	4	18 AAR30747	Substrate peptid
19	21	100.0	4	18 AAR24567	Substrate for seri
20	21	100.0	4	18 AAR08165	Proteinase site of
21	21	100.0	4	18 AAR12810	Synthetic substrat
22	21	100.0	4	19 AAR79700	B. subtilis subtil
23	21	100.0	4	19 AAR76694	B. subtilis subtil
24	21	100.0	4	19 AAR61357	Synthetic oligopep
25	21	100.0	4	19 AAR51451	Indicator for dete
26	21	100.0	4	19 AAR51610	Peptide conjugated
27	21	100.0	4	20 AAR48424	Alkaline protease
28	21	100.0	4	20 AAR07072	Alkaline protease
29	21	100.0	4	20 AAR84189	Peptide comprising
30	21	100.0	4	21 AAR20774	Chymotrypsin enzym
31	21	100.0	4	21 AAR03092	Substrate peptide
32	21	100.0	4	21 AAR01922	Synthetic pPase s
33	21	100.0	4	21 AAR91025	Model substrate pe
34	21	100.0	4	21 AAR97813	Protease peptide s
35	21	100.0	4	21 AAR80465	Chymase inhibitor
36	21	100.0	4	21 AAR76808	Chymotrypsin subst
37	21	100.0	4	21 AAR78789	Subtilisin substra
38	21	100.0	4	21 AAR49452	Cathepsin G peptid
39	21	100.0	4	21 AAR59632	Peptidylprolyl cis
40	21	100.0	4	22 AAU07699	Ztryp3 substrate u
41	21	100.0	4	22 AAR82792	Peptide substrate
42	21	100.0	4	22 AAG64601	Model peptide subs
43	21	100.0	4	22 AAG64496	Artificial substra
44	21	100.0	4	22 AAR04142	Human peptide #2
45	21	100.0	4	22 AAR598623	Peptide #2. Unde

#### ALIGNMENTS

RESULT 1

AAP40459

ID AAP40459 standard; Protein; 4 AA.

XX AAP40459;

AC

XX 27-NOV-1991 (first entry)

DT N-terminal polyhydroxyalkanoyl peptide.

DE

XX Polyhydroxyalkanoyl peptide; protease; prodrug.

XX EPI26685-A.

XX 28-NOV-1985.

XX 15-MAY-1984; 84EP-0400984.

XX 16-MAY-1983; 83FR-0308051.

XX (CNRS ) CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE.

XX Monsigny M, Mayer R;

XX WPI; 1984-296065/48.

XX N-terminal polyhydroxyalkanoyl peptide and aminoacid derivs. - opt.

XX having C-terminal active gps., e.g. chloroquine, are water-soluble

XX antitumour or antiparasitic cpds. and protease targets.

XX Claim 11; page 15; 20pp; french.

XX The peptide is polyhydroxyalkanoyl-L-Ala-L-Ala-L-Pro-L-Phe-X. X is

CC R2-NH- derived from an aromatic amine, eg p-nitro-aniline, alpha-  
 CC or beta-naphthylamine, or it is R3NH derived from an amino drug, eg  
 CC daunorubicin or chloroquine, or it is a gp. which may confer, eg  
 CC activity, eh H, OH, CH2Cl. When X is the residue of a drug, the  
 CC cpd. is a prodrug in which the peptide is the substrate for a  
 CC specific protease secreted by the target cell for the active drug,  
 CC eg tumour cells or microorganism pathogens, eg Plasmodium  
 CC falciparum. Due to the polyhydroxy gp., the cpds. are very soluble  
 CC in water or buffer soln. and are able to be used at high concn. The  
 CC cpds. also allow the detection of proteases and peptidases and  
 CC allow easy determination of the best substrate for a particular  
 CC protease.  
 CC Sequence 4 AA;  
 SQ

Query Match 100.0%; Score 21; DB 5; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAPF 4  
 DB 1 AAPF 4

RESULT 2  
 AAR04012  
 ID AAR04012 standard; peptide; 4 AA.

XX AAR04012;  
 AC AAR04012;  
 DT 19-NOV-1991 (first entry)  
 XX  
 DE Peptidase substrate analogues having peptidase inhibition activity.  
 XX  
 XX Elastase; plasminogen activator; cathepsin; calpain; enkephalinase;  
 KW Angiotensin converting enzyme.  
 XX  
 OS Synthetic.

PH Key Location/Qualifiers  
 FT Modified-site 4  
 FT /label= Terminal modified from -COOH to -COCOR  
 FT /note= "R = alkyl group or -H"

XX EP363284-A.  
 XX 11-APR-1990.  
 XX 06-OCT-1989; 89EP-0402762.  
 XX 07-OCT-1988; 88US-0254762.  
 PR 06-OCT-1989; 89EP-0402763.  
 XX (RICH ) MERRELL DOW PHARM INC.  
 PA  
 XX Bey P, Angelastro M, Mehdi S;  
 PI WPI; 1990-109579/15.  
 XX  
 DR New peptidase substrate analogue cpds. - useful as protease  
 XX inhibitors in treatment of disease states.  
 XX  
 PS Claim 5; Page 26; 33pp; English.

CC The analogues may be useful in treatment of a variety of disease  
 CC states. The scissile amide group is replaced with H or a substituted  
 CC Carbon moiety effectively inhibiting the activity of peptidases such  
 CC as elastase, plasmin thrombin, urokinase etc.  
 XX  
 XX Sequence 4 AA;

Query Match 100.0%; Score 21; DB 11; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAPF 4  
 DB 1 AAPF 4

RESULT 3  
 AAR10887  
 ID AAR10887 standard; Protein; 4 AA.

XX AAR10887;  
 AC AAR10887;  
 DT 10-APR-1991 (first entry)  
 XX  
 DE Peptide component of pentafluoroethylcarbonyl analogue.  
 XX  
 XX Protease inhibitor; rheumatoid arthritis; thrombosis; psoriasis;  
 KW male contraceptive.

XX OS Synthetic.  
 XX PN EP410411-A.  
 XX PD 30-JAN-1991.

XX 25-JUL-1990; 90EP-0114250.  
 XX 26-JUL-1989; 89US-0385624.  
 PR (RICH ) MERRELL DOW PHARM INC.  
 XX  
 PA Bey P, Peet NP, Angelastro MR, Mehdi S;  
 PI WPI; 1991-030811/05.

XX Novel serine-, carboxylic acid- and metallo-proteinase-inhibitors  
 PT - inhibit range of proteinase(s) in treating rheumatoid  
 PT arthritis, thrombosis and psoriasis, also is a male contraceptive  
 PS Claim 5; page 25; 40pp; English.  
 XX  
 CC This peptide is the R1 gp. of the cpd. of formula: R1NHCH(R2)-  
 CC COCF2CF3. This cpd. is a peptidase substrate analogue and is a  
 CC specific enzyme inhibitor for a range of proteases, e.g. serine-,  
 CC carboxylic acid-, and metallo-proteases. It is useful in the  
 CC treatment of rheumatoid arthritis, thrombosis and psoriasis and is  
 CC also used as a male contraceptive. See also AAR10876-83, AAR10886 and  
 CC AAR10888.

SQ Sequence 4 AA;

Query Match 100.0%; Score 21; DB 12; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAPF 4  
 DB 1 AAPF 4

RESULT 4  
 AAR11223  
 ID AAR11223 standard; Protein; 4 AA.

XX AAR11223;  
 AC AAR11223;  
 DT 24-MAY-1991 (first entry)  
 XX  
 DE Ketone analogue protease inhibitor #2.

XX protease inhibitor; antiinflammatory agent; hypotensive; analgesic;  
 KW antiproliferative agent; antidiemylinating agent; antithrombotic.

```

OS Synthetic.
PH Key Location/Qualifiers
FT Modified-site 1
FT Modified-site 4 /label= Methoxysuccinyl-Ala
FT Modified-site 4 /label= Phe-carboxyoxalyl
XX
XX EP417721-A.
XX
XX 20-MAR-1991.
XX
XX 11-SEP-1990; 90EP-0117461.
XX
XX 11-SEP-1989; 89US-0405491.
XX
XX (RICH ) MERRELL DOW PHARM INC.
XX
XX Flynn GA, Bey P;
XX
XX WPI; 1991-081980/12..
XX
XX New ketone analogue peptidase and isomerase inhibitors - for
PT inhibition of leukocyte elastase, cathepsin G, thrombin,
PT chymotrypsin, plasmin etc.
XX
XX Claim 5; Page 26; 50pp; English.
XX
XX This peptide is a specific example of a highly generic protease
CC inhibitor useful for medical purposes. The peptide analogues include
CC inhibitors of urokinase, renin, cathepsin D, etc. which can be used
CC as anti-proliferative agents and abortifacients, hypotensives,
CC antiinflammatory and antitumour agents, respectively.
CC See also AAR11222 and AAR11224-R11238.
XX
XX SQ Sequence 4 AA;
XX
Query Match 100.0%; Score 21; DB 12; Length 4;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4
DB 1 AAPF 4

RESULT 5
AAR29964
ID AAR29964 standard; peptide; 4 AA.
XX
AC AAR29964;
XX
DT 19-APR-1993 (first entry)
XX
DE Cathepsin G inhibiting fragment.
XX
KW Cathepsin G; elastase; connective tissue; degradation; protease;
KW gout; rheumatoid arthritis; emphysema; ARDS;
KW adult respiratory distress syndrome; para-phenylene.
XX
OS Synthetic.
XX
XX Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal amino acids of the peptides of
FT AAR29963-64 are pref. linked by
FT -C(O)-phenylene-C(O)-, esp. wherein the
FT phenylene is a para-phenylene gp."
FT Modified-site 4 /note= "C-terminal Phe is in keto form, i.e. OH
FT replaced by CGIM, pref. CF3 or CF2CF3"
PR WO9220357-A.
XX

```

```

XX
XX 26-NOV-1992.
XX
XX 21-APR-1992; 92WO-US03288.
XX
XX 23-MAY-1991; 91US-0704449.
XX
XX (RICH ) MERRELL DOW PHARM INC.
XX
XX Angelastro MR, Bey P, Doherty NS, Janusz MJ, Mehdi S;
XX Peet NP;
XX
XX WPI; 1992-415461/50.
XX
XX New peptide devlxs. used as cathepsin G and elastase inhibitors -
PT for treating gout, rheumatoid arthritis, inflammatory disorders,
PT emphysema and adult respiratory distress syndrome
XX
XX Claim 13-17; Page 52; 53pp; English.
XX
XX This sequence is an example of a highly generic formula.
CC Inhibitors of cathepsin G and elastase for preventing connective
CC tissue degradation are chemically linked inhibitors of the proteases
CC elastase (pref. the peptide of AAR29963, or Lys(2CBz)-Pro-Val or
CC Val-Pro-Val) and cathepsin G (pref. the peptide of AAR29964, or
CC Val-Pro-Phe or Phe).
CC The N-terminal amino acids of the peptides of AAR29963-64 are pref.
CC linked by -C(O)-phenylene-C(O)-, esp. wherein the phenylene is a
CC para-phenylene gp.
CC The cpds. have an anti-inflammatory effect useful in the treatment
CC of gout, rheumatoid arthritis and other inflammatory diseases and
CC to prevent elastin mediated tissue damage. They can also be used
CC in the treatment of emphysema and adult respiratory distress syndrome.
XX
XX SQ Sequence 4 AA;
XX
Query Match 100.0%; Score 21; DB 13; Length 4;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4
DB 1 AAPF 4

RESULT 6
AAR34258
ID AAR34258 standard; peptide; 4 AA.
XX
AC AAR34258;
XX
DT 19-AUG-1993 (first entry)
XX
DE Chromogenic peptide substrate for peptidase assay.
XX
KW Alkaline protease; detergent; stability.
XX
OS Synthetic.
XX
XX Key Location/Qualifiers
FT Modified-site 1 /note= "N-succinyl alanine"
FT Modified-site 4 /note= "Phe-p-nitroanilide"
XX
XX WO9307276-A.
XX
XX 15-APR-1993.
XX
XX 07-OCT-1992; 92WO-US08341.
XX
XX 08-OCT-1991; 91US-0772087.
XX

```

PA (CHEM-) CHEMGEN CORP.  
 PA (VIST-) VISTA CHEMICAL CO.  
 XX  
 PI Fodge DW, Hsiao H, Lalonde JJ;  
 XX  
 DR WPI; 1993-134465/16.  
 XX  
 DR Alkaline protease produced by bacillus stable in alkaline  
 PT conditions - used in detergents and bleaches to decompose  
 PT proteinaceous stains  
 XX  
 XX Disclosure; Page 8; 58pp; English.  
 XX  
 CC The peptide is a chromogenic substrate for use in a peptidase assay  
 CC to test the peptidase activity of a novel alkaline protease.  
 CC Detergent was found to interfere with the assay by reducing the  
 CC protease activity. See also AAR34259-61 and AAR34463-66.  
 XX  
 SQ Sequence 4 AA; Query Match 100.0%; Score 21; DB 14; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
 DB 1 AAPF 4

RESULT 7  
 AAR38410  
 ID AAR38410 standard; Protein; 4 AA.  
 AC AAR38410;  
 XX  
 XX 29-OCT-1993 (first entry)  
 DT  
 XX Cathepsin G inhibitor peptide #1.  
 DE  
 XX Inflammatory disease; prevention; rheumatoid arthritis; emphysema;  
 KW neutrophil-mediated connective tissue degradation; gout; elastase;  
 KW inhibition; adult respiratory distress syndrome.  
 XX  
 XX Synthetic.  
 OS  
 XX Key Location/Qualifiers  
 FT Modified-site 1 /note= "Linked via a linking chain such as  
 FT -CO-phenylene-CO- to the N-terminal of a  
 FT Cathepsin G inhibiting molecule as described  
 FT in AAR38410-11"  
 FT  
 FT Modified-site 4 /note= "May opt. be modified by COCOR, CF2CF3, CF3,  
 FT CHF2, COOR3, CONHR3, CF2CHRC3CONHR, H, alkyl,  
 FT aryl, aralkyl or COR; where R3 is H, alkyl,  
 FT phenyl or benzyl and R is OH or alkoxy"  
 XX  
 PN ZA9203602-A.  
 XX  
 XX 24-FEB-1993.  
 XX  
 XX 18-MAY-1992; 92ZA-0003602.  
 XX  
 XX 23-MAY-1991; 91US-0704499.  
 PR  
 XX (RICH ) MERRELL DOW PHARM INC.  
 PA  
 XX Angelastro MM, Bey P, Doherty NS, Janusz MJ, Mehdi S;  
 PI Peet NP;  
 PI  
 XX WPI; 1993-197380/24.  
 DR  
 XX New cathepsin G and elastase inhibitors - prevents connective  
 PT

PT tissue degradation  
 XX Claim 13-17; Page 52; 55pp; English.  
 XX  
 CC The sequences given in AAR38410-11 are cathepsin G inhibiting peptides  
 CC which were produced by standard peptide synthesis methods. In the  
 CC context of the invention one of these peptides may be linked via  
 CC their N-termini to an elastase inhibiting peptide (see features  
 CC table). The peptide conjugates may be used to prevent neutrophil-  
 CC mediated connective tissue degradation associated with inflammatory  
 CC diseases eg. gout and rheumatoid arthritis. They may also be used  
 CC for preventing elastin-mediated tissue damage in the treatment of  
 CC emphysema and adult respiratory distress syndrome.  
 XX  
 SQ Sequence 4 AA; Query Match 100.0%; Score 21; DB 14; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
 DB 1 AAPF 4

RESULT 8  
 AAR44109  
 ID AAR44109 standard; peptide; 4 AA.  
 AC AAR44109;  
 XX  
 XX 16-MAY-1994 (first entry)  
 DT  
 XX Aminonaphthalene propyl sulphamide tetrapeptide.  
 DE  
 XX Chymotrypsin assay; ansa-substrate; detector group.  
 KW  
 XX Synthetic.  
 OS  
 XX Key Location/Qualifiers  
 FT Modified-site 1 /label= OTHER  
 FT /note= "Suc-Ala (Claim 1) or Z-Ala (Claim 2)"  
 FT Modified-site 4 /label= OTHER  
 FT /note= "Phe-NH-(1,5-naphthylene)-SO2NH-C3H7"  
 FT  
 XX SU1771478-A.  
 PN  
 XX 23-OCT-1992.  
 PD  
 XX 20-JUL-1990; 90SU-4872552.  
 PF  
 XX 20-JUL-1990; 90SU-4872552.  
 PR  
 XX (ALIB=) AS LITH BIOCHEM INST.  
 XX (ASMO=) AS USSR MOLECULAR GENETICS.  
 PA  
 XX Nedospasov AA, Palaima AI, Yanchene RA;  
 PI  
 XX WPI; 1993-358242/45.  
 DR  
 XX  
 XX Prepn. of new succinyl-alanyl-alanyl-propyl-phenylalanyl-amino-  
 PT naphthalen-propyl-sulphamide - for use as ansa-substrate in the  
 PT quantitative determ. of chymotrypsin  
 PT  
 XX Claim 1 and 2; Column 8; 4pp; Russian.  
 PS  
 XX 5-(N-succinyl-alanyl-alanyl-propyl-phenylalanyl)aminonaphthalene-1-  
 CC (N-propyl) sulphamide (Claim 1) and 5-(N-benzoyloxycarbonyl-alanyl-  
 CC alanyl-propyl-phenylalanyl)aminonaphthalene-1-(N-propyl) sulphamide  
 CC (claim 2) are prepared by initially reacting N-benzoyloxycarbonyl-  
 CC phenylalanine with isobutyl chloroformate in tetrahydrofuran

```

CC followed by reaction with 5-aminonaphthalene-1-(N-propyl)sulphamide.
CC The new compounds are used as detector groups in quantitative
CC determination of chymotrypsin.
XX
XX
SQ Sequence 4 AA;

Query Match 100.0%; Score 21; DB 14; Length 4;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAPF 4
Db 1 AAPF 4

RESULT 9
AAR52024
ID AAR52024 standard; peptide; 4 AA.
XX
AC AAR52024;
XX
DT 04-AUG-1994 (first entry)
XX
DE Serine protease inhibitor tetrapeptide.
XX
KW Prevention; schistosomiasis; parasite; infection; prevention;
KW parasitic penetration; skin; cercariae; anti-penetrant.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT Modified-site 4 /note= "BG(peptide blocking gp.) attached"
FT Modified-site 4 /note= "PI(protease inhibitor), other than
FT chloromethyl ketone, attached"
XX
PN US5284829-A.
XX
PD 08-FEB-1994.
XX
PF 26-NOV-1991; 9IUS-0798565.
XX
PR 26-NOV-1991; 9IUS-0798565.
XX
PA (REGC ) UNIV CALIFORNIA.
XX
PI Cohen FE, McKerrow JH;
XX
DR WPI; 1994-056364/07.
XX
PT Synthetic tetrapeptide(s) with an N-terminal blocking gp. and
PT C-terminal enzyme inhibitor - can be formulated into soaps and
PT sprays and used to prevent schistosomal skin penetration
XX
PS Disclosure; Page 7; 35pp; English.
XX
CC The sequence is that of a synthetic tetrapeptide serine protease
CC inhibitor which can be used to prevent schistosome parasite
CC infection. It may be used in a formulation as a soap, lotion,
CC cream, spray, etc. to stop parasitic penetration of the skin.
XX
SQ Sequence 4 AA;

Query Match 100.0%; Score 21; DB 15; Length 4;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAPF 4
Db 1 AAPF 4

RESULT 11
AAR53781
ID AAR53781 standard; Peptide; 4 AA.
XX
AC AAR53781;
XX
DT 29-DEC-1994 (first entry)
XX
DE Sequence of internal fragment of a chromogenic substrate for
DE cathepsin G.
XX
KW Enzyme; cathepsin G; protease; chromogenic substrate.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /label= Suc-Ala
FT

CC followed by reaction with 5-aminonaphthalene-1-(N-propyl)sulphamide.
CC The new compounds are used as detector groups in quantitative
CC determination of chymotrypsin.
XX
XX
SQ Sequence 4 AA;

Query Match 100.0%; Score 21; DB 14; Length 4;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAPF 4
Db 1 AAPF 4

RESULT 9
AAR52024
ID AAR52024 standard; peptide; 4 AA.
XX
AC AAR52024;
XX
DT 28-NOV-1994 (first entry)
XX
DE Bacillus alkali protease substrate.
XX
KW Bacillus NKS-21 alkali protease; hydrolysis; thermostable enzyme;
KW detergent stable.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /label= OTHER
FT Modified-site 4 /note= "Suc-Ala"
FT Modified-site 4 /label= OTHER
FT Modified-site 4 /note= "Phe-MCA"
XX
PN JP06070765-A.
XX
PD 15-MAR-1994.
XX
PF 08-OCT-1992; 92JP-0296360.
XX
PR 10-JUL-1992; 92JP-0207302.
XX
PA (SHOW ) SHOWA DENKO KK.
XX
DR WPI; 1994-128672/16.
XX
PT New alkali protease stable to heat and detergent - useful as
PT industrial enzyme, eg in washing compositions
XX
PS Claim 1; Page 2; 10pp; Japanese.
XX
CC A novel alkali protease obtained from Bacillus NKS-21 is defined by
CC its physico-chemical properties including substrate specificity for
CC the tetrapeptide AAR52024.
XX
SQ Sequence 4 AA;

Query Match 100.0%; Score 21; DB 15; Length 4;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAPF 4
Db 1 AAPF 4

RESULT 10
AAR46223
ID AAR46223 standard; peptide; 4 AA.

```

```

FT Modified-site 4 /note= "Succinyl group"
FT FT /label= Phe-pNA
FT FT /note= "p-nitroanil"
XX
XX WO9412637-A.
XX
XX
XX PD 09-JUN-1994.
XX
XX PF 01-DEC-1993; 93WO-US11696.
XX
XX PR 02-DEC-1992; 92US-0985692.
XX PR 19-NOV-1993; 93US-0155331.
XX
XX PA (NOVO ) NOVO-NORDISK AS.
XX PA (ZYMO ) ZYMOGENETICS INC.
XX
XX PI Foster DC, Norris K, Sprecher CA;
XX
XX DR WPI; 1994-200265/24.
XX
XX PT New human Kunitz type protease inhibitor and related DNA - for
PT treating pancreaticis and other disorders involving serine
PT protease, also new amyloid protein precursor homologues including
PT the inhibitor in its sequence
XX
XX PS Example; Page 63; 70pp; English.
XX
XX CC Protease inhibitory profiles of the Kunitz inhibitors
CC were determined for a variety of proteases using
CC a variety of chromogenic substrates and compared to the inhibitory
CC activity shown by the Kunitz-type inhibitor domain of the amyloid
CC protein precursor and bovine aprotinin. The substrate AARS3780
CC was tested using the protease leukocyte elastase.
XX
XX SQ Sequence 4 AA;

Query Match 100.0%; Score 21; DB 15; Length 4;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4
DB |||||
1 AAPF 4

RESULT 12
AAR72920
ID AAR72920 standard; Peptide; 4 AA.
XX
XX AC AAR72920;
XX
XX DT 29-NOV-1995 (first entry)
XX
XX DE Substrate for peptidyl prolyl cis trans isomerase alpha.
XX
XX KW Escherichia coli; protein conformation; folding; acceleration;
XX KW PPIase-alpha; peptidyl prolyl cis trans isomerase alpha;
XX KW catalysis; isomerisation; prolyl peptide bond.
XX
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-succinyl-Ala"
FT FT Modified-site 4
FT FT Modified-site 4 /note= "Phe-MCA (sic)"
XX
XX PN EP647714-A.
XX
XX PD 12-APR-1995.
XX
XX PF 19-JUL-1990; 90EP-0307914.
XX
XX PR 19-JUL-1989; 89JP-0184738.
XX PR 06-OCT-1989; 89JP-0260244.
XX PR 29-DEC-1989; 89JP-0344705.
XX
XX PA (TOFU ) TONEN CORP.
XX
XX PI Hayano T, Katou S, Maki N, Suzuki M, Takahashi N;
XX
XX DR WPI; 1995-140755/19.

Query Match 100.0%; Score 21; DB 16; Length 4;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4
DB |||||
1 AAPF 4

RESULT 13
AAR72876
ID AAR72876 standard; Peptide; 4 AA.
XX
XX AC AAR72876;
XX
XX DT 29-NOV-1995 (first entry)
XX
XX DE Substrate for peptidyl prolyl cis trans isomerase alpha.
XX
XX KW Escherichia coli; protein conformation; folding; acceleration;
XX KW PPIase-alpha; peptidyl prolyl cis trans isomerase alpha;
XX KW catalysis; isomerisation; prolyl peptide bond.
XX
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-succinyl-Ala"
FT FT Modified-site 4
FT FT Modified-site 4 /note= "Phe-MCA (sic)"
XX
XX PN EP647713-A.
XX
XX PD 12-APR-1995.
XX
XX PF 19-JUL-1990; 90EP-0307914.
XX
XX PR 19-JUL-1989; 89JP-0184738.
XX PR 06-OCT-1989; 89JP-0260244.
XX PR 29-DEC-1989; 89JP-0344705.
XX
XX PA (TOFU ) TONEN CORP.
XX
XX PI Hayano T, Katou S, Maki N, Suzuki M, Takahashi N;
XX
XX DR WPI; 1995-140755/19.

```

XX New E.coli peptidyl prolyl cis trans isomerase alpha - used to  
PT accelerate the folding of proteins, partic. for activation of  
PT inactive recombinant proteins  
XX  
PS Disclosure; Page 5; 85pp; English.  
XX  
CC To measure the activity of E. coli PPIase-alpha (peptidyl prolyl cis  
CC trans isomerase alpha), the PPIase and Csa (cyclosporine A) are  
CC added in the cell and mixed for 1 minute. Then, this peptide is  
CC added and incubated, and chymotrypsin added to start the reaction.  
CC The effect of Csa on the inhibition of the PPIase activity can be  
CC detected by varying the amt. of the Csa added. The inventors are  
CC claiming the PPIase-alpha.  
XX  
XX Sequence 4 AA;  
SQ  
Query Match 100.0%; Score 21; DB 16; Length 4;  
Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AAPF 4  
DB 1 AAPF 4  
RESULT 14  
ID AAR77315 standard; Protein; 4 AA.  
AC AAR77315;  
XX  
XX 28-FEB-1996 (first entry)  
XX  
XX Porphyromonas gingivalis protease substrate.  
XX  
XX Protease; periodontal disease; pathogenic microbe; diagnosis;  
XX substrate.  
XX  
XX Synthetic.  
XX  
XX Key Location/Qualifiers  
FH Modified-site 1 /label= "Suc-Ala"  
FT Modified-site 4 /note= "MCA-Phe"  
FT  
XX JP07135973-A.  
XX  
XX 30-MAY-1995.  
XX  
XX 15-NOV-1993; 93JP-0307084.  
XX  
XX 15-NOV-1993; 93JP-0307084.  
XX (SUNR ) SUNTORY LTD.  
XX  
XX WPI; 1995-227397/30.  
XX  
XX An enzyme originated from a periodontal disease pathogenic microbe -  
XX and an antibody against the enzyme, for the determination of the  
XX progress and activity of the disease  
XX  
XX Example; Page 7; 15pp; Japanese.  
XX  
XX AAR77315 is a target substrate for a Porphyromonas gingivalis (a  
XX periodontal disease pathogenic microbe) protease. An antibody  
XX raised against the enzyme can be used to diagnose the presence  
XX and progress of a periodontal disease, caused by a pathogenic  
XX microbe.  
XX  
XX Sequence 4 AA;

Query Match 100.0%; Score 21; DB 16; Length 4;  
Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AAPF 4  
DB 1 AAPF 4  
RESULT 15  
ID AAR77196 standard; peptide; 4 AA.  
XX  
XX AAR77196;  
XX  
XX 27-FEB-1996 (first entry)  
XX  
XX Cell proliferation enzyme proteinase activity substrate peptide #1.  
XX  
XX Proteinase; cell growth-stimulating protein; hydrolysis;  
XX macrophage chemotactic action; serine protease inhibitor; wound;  
XX gastric ulcer; leg ulcer; bed sore.  
XX  
XX Synthetic.  
XX  
XX Key Location/Qualifiers  
FH Modified-site 1 /label= Succinyl-Ala  
FT Modified-site 4 /note= "Modified by 4-methyl-coumaryl-7-amide"  
FT  
XX EP661293-A2.  
XX  
XX 05-JUL-1995.  
XX  
XX 22-DEC-1994; 94EP-0120406.  
XX  
XX 29-DEC-1993; 93JP-0351225.  
XX  
XX (SANW ) SANWA KAGAKU KENKYUSHO CO.  
XX  
XX Awaya J, Uesaka H, Watanabe K, Yamaguchi T;  
XX WPI; 1995-233274/31.  
XX  
XX Protein derived from Clostridium perfringens FERM BP-4584 -  
XX stimulates cell proliferation and has macrophage chemotactic action  
XX  
XX Example; Page 8; 14pp; English.  
XX  
XX The sequences given in AAR77196-200 are peptides which were used to  
XX demonstrate the proteinase activity of the cell growth-stimulating  
XX protein of the invention. The protein was seen to selectively  
XX hydrolyse synthetic substrates having an aromatic amino acid at the  
XX C-terminal. The response to the peptide given in AAR77196 was  
XX particularly high. The cell growth-stimulating protein has a mol.  
XX wt. of 420 +/- 40 kD and a single subunit mol. wt. of 130 +/- 20 kD.  
XX It has an isoelectric point of 4.8 and has cell growth stimulating  
XX action and macrophage chemotactic action, as well as proteinase  
XX activity. Its enzymatic activity decreases in the presence of a  
XX serine protease inhibitor, increase in the presence of various metal  
XX ions, and is stabilised in the presence of calcium ion. The protein is  
XX particularly useful for the treatment of wounds, gastric and leg ulcers,  
XX eg. bed sores.  
XX  
XX Sequence 4 AA;  
SQ  
Query Match 100.0%; Score 21; DB 16; Length 4;  
Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AAPF 4  
DB 1 AAPF 4

Db 1 AAPF 4

Search completed: February 12, 2003, 10:22:28  
Job time : 4.68657 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:17:50 ; Search time 1.00299 Seconds  
(without alignments)  
383.393 Million cell updates/sec

Title: US-10-036-371-7

Perfect score: 21

Sequence: 1 AAPF 4

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 73:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	5	2 JS0319	subesophageal gang
2	21	100.0	42	2 S43963	hypothetical prote
3	21	100.0	47	2 PN0607	cytochrome-c oxida
4	21	100.0	50	2 AD3575	hypothetical prote
5	21	100.0	63	1 OSB08A	cytochrome-c oxida
6	21	100.0	63	2 S10303	hypothetical prote
7	21	100.0	75	2 AB2778	hypothetical prote
8	21	100.0	79	2 C69125	hypothetical prote
9	21	100.0	80	2 G96953	probable metal-bin
10	21	100.0	86	2 E90905	hypothetical prote
11	21	100.0	88	2 F97429	filio protein (AF30
12	21	100.0	88	2 AG2647	flagellar biosynth
13	21	100.0	90	2 S02854	male accessory gla
14	21	100.0	90	2 S30428	Met26Ab protein -
15	21	100.0	90	2 S30412	male accessory gla
16	21	100.0	90	2 S30430	Met26Ab protein -
17	21	100.0	94	2 C98239	hypothetical prote
18	21	100.0	97	2 AB1502	probable cell surf
19	21	100.0	100	2 G86174	hypothetical prote
20	21	100.0	101	2 D90761	hypothetical prote
21	21	100.0	102	2 B86976	PF-family protein
22	21	100.0	102	2 H70898	probable EG protei
23	21	100.0	107	2 H75397	hypothetical prote
24	21	100.0	108	2 F72507	hypothetical prote
25	21	100.0	109	2 G86433	protein T17H7.7 [i
26	21	100.0	112	2 H96031	probable transcrip
27	21	100.0	122	2 PC4279	anti-SS-A/Ro 60K p
28	21	100.0	122	2 PC4280	anti-SS-A/Ro 60K p
29	21	100.0	122	2 CB3501	hypothetical prote

30	21	100.0	125	2 AD0605	probable membrane
31	21	100.0	125	2 A83048	hypothetical prote
32	21	100.0	126	2 T05908	probable ribosomal
33	21	100.0	127	2 A25864	calcitonin gene-re
34	21	100.0	127	2 A71334	hypothetical prote
35	21	100.0	128	1 TCHUR	calcitonin gene-re
36	21	100.0	129	2 137779	Ig variable region
37	21	100.0	130	2 T29498	hypothetical prote
38	21	100.0	132	2 S36196	hypothetical prote
39	21	100.0	133	2 S76285	hypothetical prote
40	21	100.0	133	2 F75297	hypothetical prote
41	21	100.0	133	2 S75430	calcitonin gene-re
42	21	100.0	134	2 A44173	calcitonin gene-re
43	21	100.0	134	2 A72489	hypothetical prote
44	21	100.0	137	2 F87627	conserved hypothet
45	21	100.0	139	2 D75542	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

JS0319

subesophageal ganglion pentapeptide - house cricket

C;Species: Acheta domesticus (house cricket)

C;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 20-Jun-2000

C;Accession: JS0319

R;Wicker, C.; Wicker, C.

Comp. Biochem. Physiol. C 88, 185-187, 1987

A;Title: Isolation and structure of a peptide isolated from the subesophageal ganglion (

A;Reference number: JS0319

A;Accession: JS0319

A;Molecule type: protein

A;Residues: 1-5 <MIC>

Query Match 100.0%; Score 21; DB 2; Length 5;

Best Local Similarity 100.0%; Pred. No. 2.8e+05;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4

DB 2 AAPF 5

##### RESULT 2

S43963

hypothetical protein (clone pR57) - Rhizobium sp. (strain NGR234) (fragment)

C;Species: Rhizobium sp.

A;Variety: strain NGR234

C;Date: 20-Oct-1994 #sequence\_revision 23-Feb-1996 #text\_change 21-Aug-1998

C;Accession: S43963

R;Perret, X.; Fellay, R.; Boursion, A.J.; Cooper, J.E.; Brenner, S.; Broughton, W.J.

Nucleic Acids Res. 22, 1335-1341, 1994

A;Title: Subtraction hybridisation and shot-gun sequencing: a new approach to identify s

A;Reference number: S43961; MUID:94248027; PMID:8190622

A;Accession: S43963

A;Status: nucleic acid sequence not shown

A;Molecule type: DNA

A;Residues: 1-42 <PER>

A;Experimental source: strain NGR234

C;Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 100.0%; Score 21; DB 2; Length 42;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4

DB 10 AAPF 13

##### RESULT 3

PN0607

cytochrome-c oxidase (EC 1.9.3.1) chain VIIC - pig (fragment)  
 C;Species: Sus scrofa domestica (domestic pig)  
 C;Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 07-May-1999  
 C;Accession: PN0607  
 R;Sillard, R.; Joenwall, H.; Mutt, V.  
 Biochem. Biophys. Res. Commun. 195, 746-750, 1993  
 A;Title: Characterization of porcine intestinal cytochrome c oxidase subunit VIIC, purified from porcine small intestine  
 A;Reference number: PN0607; MUID:93384597; PMID:8396926  
 A;Accession: PN0607  
 A;Molecule type: protein  
 A;Residues: 1-47 <SIL>  
 A;Experimental source: intestine  
 A;Superfamily: cytochrome-c oxidase chain VIIC  
 C;Keywords: mitochondrion; oxidoreductase; respiratory chain

Query Match 100.0%; Score 21; DB 2; Length 47;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPP 4  
 ||||  
 Db 34 AAPF 37

RESULT 4  
 AD3575  
 hypothetical protein BMEI10525 [imported] - Brucella melitensis (strain 16M)  
 C;Species: Brucella melitensis  
 C;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
 C;Accession: AD3575  
 R;DeVechio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujex, C.; Los, T.; Ivanova,  
 ; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letes  
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
 A;Reference number: AD3575; PMID:1175668  
 A;Accession: AD3575  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-50 <KUR>  
 A;Cross-references: GB:AE008918; PIDN:AAU53767.1; PID:gl7984695; GSPDB:GN00191  
 A;Experimental source: strain 16M  
 C;Genetics:  
 A;Gene: BMEI10525  
 A;Map position: II

Query Match 100.0%; Score 21; DB 2; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPP 4  
 ||||  
 Db 44 AAPF 47

RESULT 5  
 OSB08A  
 cytochrome-c oxidase (EC 1.9.3.1) chain VIIC precursor [validated] - bovine  
 N;Alternate names: cytochrome-c oxidase chain VIIIA  
 C;Species: Bos primigenius taurus (cattle)  
 C;Date: 31-May-1979 #sequence\_revision 31-Dec-1992 #text\_change 15-Sep-2000  
 C;Accession: JH0473; S06597; A00498; H29968; S18834  
 R;Aqua, M.S.; Bachman, N.J.; Lomax, M.I.; Grossman, L.I.  
 Gene 104, 211-217, 1991  
 A;Title: Characterization and expression of a cDNA specifying subunit VIIC of bovine cytochrome c oxidase  
 A;Reference number: JH0473; MUID:92009215; PMID:1655579  
 A;Accession: JH0473  
 A;Molecule type: DNA  
 A;Residues: 1-63 <AQU1>  
 A;Cross-references: GB:X58823  
 R;Aqua, M.S.; Lomax, M.I.; Schon, E.A.; Grossman, L.I.  
 Nucleic Acids Res. 17, 8376, 1989  
 A;Title: Nucleotide sequence of a cDNA for bovine cytochrome c oxidase subunit VIIC.  
 A;Reference number: S06597; MUID:90045968; PMID:2554257

A;Accession: S06597  
 A;Molecule type: mRNA  
 A;Residues: 1-63 <AQU2>  
 A;Cross-references: EMBL:X15725; NID:9277; PIDN:CAB57793.1; PID:G6015488  
 R;Buse, G.; Steffens, G.J. Chem. 359, 1005-1009, 1978  
 Hoppe-Seyler's Z. Physiol. Chem. 359, 1005-1009, 1978  
 A;Title: Studies on cytochrome c oxidase, II. The chemical constitution of a short polypeptide chain from bovine heart  
 A;Reference number: A00498; MUID:79046803; PMID:213363  
 A;Accession: A00498  
 A;Molecule type: protein  
 A;Residues: 17-63 <BUS>  
 A;Experimental source: heart  
 R;Yanamura, W.; Zhang, Y.Z.; Takamiya, S.; Capaldi, R.A.  
 Biochemistry 27, 4909-4914, 1988  
 A;Title: Tissue-specific differences between heart and liver cytochrome c oxidase.  
 A;Reference number: A90531; MUID:89000697; PMID:2844245  
 A;Accession: H29968  
 A;Molecule type: protein  
 A;Residues: 17-42 <YAN>  
 A;Experimental source: liver  
 R;Tsukihara, T.; Aoyama, H.; Yamashita, E.; Tomizaki, T.; Yamaguchi, H.; Shinzawa-Itch, F.  
 submitted to the Brookhaven Protein Data Bank, April 1996  
 A;Reference number: A67451; PDB:1OCC  
 A;Content: annotation; X-ray crystallography, 2.8 angstroms, residues 17-63  
 R;Tsukihara, T.; Aoyama, H.; Yamashita, E.; Tomizaki, T.; Yamaguchi, H.; Shinzawa-Itch, K.  
 Science 272, 1136-1144, 1996  
 A;Title: The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 angstrom resolution  
 A;Reference number: A57981; MUID:96216288; PMID:8638158  
 A;Content: annotation; X-ray crystallography, 2.8 angstroms  
 C;Genetics:  
 A;Genome: nuclear  
 C;Complex: part of a 13 chain complex spanning the inner mitochondrial membrane and consisting of four molecules of cytochrome c oxidase and two molecules of cytochrome b (see PIR:CA80), Vb (see PIR:OGB06), Vlb (see PIR:OGB07), Vlc (see PIR:OGB08) and two molecules of cytochrome c oxidase (see PIR:OGB09)  
 C;Function:  
 A;Description: the cytochrome-c oxidase complex catalyzes the oxidation of four molecules of reduced cytochrome c to oxidized cytochrome c, producing two molecules of water and lowering the concentration of cytochrome c in the mitochondrial matrix  
 A;Pathway: oxidative phosphorylation; respiratory chain  
 A;Note: the role of chain VIIC is not clear  
 C;Superfamily: cytochrome-c oxidase chain VIIC  
 F;1-16/Domain: electron transfer; membrane-associated complex; mitochondrial inner membrane, F;1-16/Domain: transit peptide (mitochondrion) #status predicted <TNP>  
 F;17-63/Product: cytochrome-c oxidase chain VIIC #status experimental <MAT>  
 F;17-33/Domain: mitochondrial matrix #status experimental <MM1>  
 F;34-60/Domain: transmembrane helix #status experimental <TR01>  
 F;61-63/Domain: intracristal #status experimental <ITC1>

Query Match 100.0%; Score 21; DB 1; Length 63;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
 ||||  
 Db 50 AAPF 53

RESULT 6  
 S10303  
 cytochrome-c oxidase (EC 1.9.3.1) chain VIIC - mouse  
 C;Species: Mus musculus (house mouse)  
 C;Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 11-Jun-1999  
 C;Accession: S10303  
 R;Akamatsu, M.; Grossman, L.I.  
 Nucleic Acids Res. 18, 3645, 1990  
 A;Title: Nucleotide sequence of a cDNA for mouse cytochrome c oxidase subunit VIIC.  
 A;Reference number: S10303; MUID:90301494; PMID:2163523  
 A;Accession: S10303  
 A;Molecule type: mRNA  
 A;Residues: 1-63 <AKA>  
 A;Cross-references: EMBL:X52940; NID:950524; PIDN:CAA37115.1; PID:950525  
 C;Genetics:  
 A;Genome: nuclear  
 C;Superfamily: cytochrome-c oxidase chain VIIC

C;Keywords: mitochondrion; oxidoreductase

```

Query Match      100.0%; Score 21; DB 2; Length 63;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4
    ||||
Db 50 AAPF 53

RESULT 7
AB2778
hypothetical protein Atu1639 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C;Accession: AB2778
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
Rage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Author: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AB2778
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-75 <KUR>
A;Cross-references: GB:AE008698; PIDN:AA42640.1; PID:g17740071; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu1639
A;Map position: circular chromosome

Query Match      100.0%; Score 21; DB 2; Length 75;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4
    ||||
Db 50 AAPF 53

RESULT 8
C69125
hypothetical protein MTH206 - Methanobacterium thermoautotrophicum (strain Delta H)
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C;Accession: C69125
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jilwani, N.
; K.; S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: C69125
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-79 <MTH>
A;Cross-references: GB:AE000807; GB:AE000666; NID:g2621239; PIDN:AA884712.1; PID:g262125
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH206
A;Start codon: TTG

Query Match      100.0%; Score 21; DB 2; Length 79;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4
    ||||
Db 36 AAPF 39

RESULT 9
G96953
probable metal-binding protein [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: G96953
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, O.; Gibson, R.; Lee
; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: G96953
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-80 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK78418.1; PID:g15023293; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC0438

Query Match      100.0%; Score 21; DB 2; Length 80;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4
    ||||
Db 63 AAPF 66

RESULT 10
E90905
hypothetical protein ECS2213 [imported] - Escherichia coli (strain O157:H7, substrain R1)
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: E90905
R;Hayashi, T.; Makino, K.; Kurokawa, K.; Iehii, K.; Yokoyama, K.; Han, C.G.
gawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genor
A;Reference number: A95629; MUID:21156231; PMID:11258796
A;Accession: E90905
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-86 <HAY>
A;Cross-references: GB:BA000007; PIDN:BA835636.1; PID:g13361679; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain R1MD 0509952
C;Genetics:
A;Gene: ECS2213

Query Match      100.0%; Score 21; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4
    ||||
Db 26 AAPF 29

RESULT 11
F97429
fliQ protein (AF300968) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C;Accession: F97429
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman
; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: F97429
A;Status: preliminary

```

A;Molecule type: DNA  
A;Residues: 1-88 <KUR>  
A;Cross-references: GB:AE007869; PIDN:AAK86391.1; PID:gl15155521; GSPDB:GN00169  
C;Genetics:  
A;Gene: AGR\_C\_1023  
A;Map position: circular chromosome

Query Match 100.0%; Score 21; DB 2; Length 88;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAPF 4  
Db 65 AAPF 68

RESULT 12  
AG2647  
flagellar biosynthetic protein [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C;Species: Agrobacterium tumefaciens  
C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C;Accession: AG2647  
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Erge, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan, Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.W.  
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A;Reference number: AB2577; PMID:11743193  
A;Accession: AG2647  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-88 <KUR>  
A;Cross-references: GB:AE008688; PIDN:AAL41597.1; PID:gl17738933; GSPDB:GN00186  
A;Experimental source: strain C58 (Dupont)  
C;Genetics:  
A;Gene: fliQ  
A;Map position: circular chromosome

Query Match 100.0%; Score 21; DB 2; Length 88;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAPF 4  
Db 65 AAPF 68

RESULT 13  
S02854  
male accessory gland secretory protein 26Ab precursor - fruit fly (Drosophila melanogaster)  
N;Alternate names: male accessory gland secretory protein msp355b; male paragonial protein  
C;Species: Drosophila melanogaster  
C;Date: 01-Dec-1989 #sequence\_revision 01-Dec-1989 #text\_change 21-Jul-2000  
A;Accession: S02854; S30408; S30416; S30420; S30426; S30410; S30424  
R;Monma, S.A.; Wolfner, M.F.  
Genes Dev. 2, 1063-1073, 1988  
A;Title: Structure and expression of a Drosophila male accessory gland gene whose product is required for male fertility  
A;Reference number: S02853; MUID:89053045; PMID:3142802  
A;Accession: S02854  
A;Molecule type: DNA  
A;Residues: 1-90 <MON>  
A;Cross-references: EMBL:Y00219; NID:g8264; PIDN:CAA68367.1; PID:g8266  
A;Experimental source: strain Canton-S  
R;Aguade, M.; Miyashita, N.; Langley, C.H.  
Genetics 132, 755-770, 1992  
A;Title: Polymorphism and divergence in the Mat26A male accessory gland gene region in Drosophila  
A;Reference number: S30407; MUID:93106377; PMID:1361475  
A;Accession: S30408  
A;Molecule type: DNA  
A;Residues: 1-90 <AGU>  
A;Cross-references: EMBL:X70888; NID:g3402845; PIDN:CAA50233.1; PID:g8234

A;Experimental source: allele NC1  
A;Accession: S30416  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-90 <AGW>  
A;Cross-references: EMBL:X70892; NID:g8244; PID:g8246  
A;Experimental source: allele NC5  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
A;Accession: S30420  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-90 <AGF>  
A;Cross-references: EMBL:X70894; NID:g8250; PID:g8252  
A;Experimental source: allele NC7  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
A;Accession: S30426  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-90 <AGA>  
A;Cross-references: EMBL:X70897; NID:g8259; PID:g8260  
A;Experimental source: allele NC10  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
A;Accession: S30410  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-79, 'V', '81-90 <AG3>  
A;Cross-references: EMBL:X70889; NID:g3402846; PIDN:CAA50235.1; PID:g8237  
A;Experimental source: allele NC2  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
A;Accession: S30424  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-63, 'H', '65-90 <AG4>  
A;Cross-references: EMBL:X70896; NID:g3406838; PIDN:CAA50249.1; PID:g8258  
A;Experimental source: allele NC9  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
C;Genetics:  
A;Gene: FlyBase:Acp26Ab  
A;Cross-references: FlyBase:FBgn0002856  
A;Map position: 2  
A;Introns: 11/1  
C;Superfamily: male accessory gland secretory protein 26Ab  
F;1-18/Domain: signal sequence #status predicted <SIG>  
F;19-90/Product: male accessory gland secretory protein 26Ab #status predicted <MAT>

Query Match 100.0%; Score 21; DB 2; Length 90;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAPF 4  
Db 21 AAPF 24

RESULT 14  
S30428  
Mat26Ab protein - fruit fly (Drosophila mauritiana)  
C;Species: Drosophila mauritiana  
C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 21-Jul-2000  
R;Aguade, M.; Miyashita, N.; Langley, C.H.  
Genetics 132, 755-770, 1992  
A;Title: Polymorphism and divergence in the Mat26A male accessory gland gene region in Drosophila  
A;Reference number: S30407; MUID:93106377; PMID:1361475  
A;Accession: S30428  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-90 <AGU>  
A;Cross-references: EMBL:X70898; NID:g8261; PIDN:CAA50253.1; PID:g8263  
C;Genetics:  
A;Gene: FlyBase:Dmau/Acp26Ab  
A;Cross-references: FlyBase:FBgn0012495  
A;Introns: 11/1

C:Superfamily: male accessory gland secretory protein 26Ab

Query Match 100.0%; Score 21; DB 2; Length 90;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AAPF 4  
|||  
Db 21 AAPF 24

# RESULT 15

S30412  
male accessory gland secretory protein 26Ab precursor (variant 1) - fruit fly (Drosophila)  
N:Alternate names: male accessory gland secretory protein 355b; Mat26Ab protein  
C:Species: Drosophila melanogaster  
C:Date: 02-Dec-1993 #sequence revision 26-May-1995 #text\_change 21-Jul-2000  
C:Accession: S30412; S30414; S30418; S30422  
R:Aguiade, M.; Miyashita, N.; Langley, C.H.  
Genetics 132, 755-770, 1992  
A:Title: Polymorphism and divergence in the Mat26A male accessory gland gene region in D.  
A:Reference number: S30407; MUID:93106377; PMID:1361475  
A:Accession: S30412  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-90 <AGU>  
A:Cross-references: EMBL:X70890; NID:93402847; PIDN:CAA50237.1; PID:g8240  
A:Experimental source: allele NC3  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
A:Accession: S30414  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-90 <AGW>  
A:Cross-references: EMBL:X70891; NID:g8241; PID:g8243  
A:Experimental source: allele NC4  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
A:Accession: S30418  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-90 <AGF>  
A:Cross-references: EMBL:X70893; NID:g8247; PID:g8249  
A:Experimental source: allele NC6  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
A:Accession: S30422  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-90 <AGA>  
A:Cross-references: EMBL:X70895; NID:g8253; PID:g8255  
A:Experimental source: allele NC8  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
C:Genetics:  
A:Gene: FlyBase:Acp26Aa  
A:Cross-references: FlyBase:FBgn0002855  
A:Map position: 2  
A:Introns: 11/1  
C:Superfamily: male accessory gland secretory protein 26Ab  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-90/Product: male accessory gland secretory protein 26Ab #status predicted <MAT>

Query Match 100.0%; Score 21; DB 2; Length 90;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AAPF 4  
|||  
Db 21 AAPF 24

Search completed: February 12, 2003, 10:28:55  
Job time : 4.00299 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:04:45 ; Search time 0.537313 Seconds  
(without alignments)  
308.768 Million cell updates/sec

Title: US-10-036-371-7

Perfect score: 21

Sequence: 1 AAPF 4

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	5	1 SUGA_ACHDO	P19991 acheta dome
2	21	100.0	63	1 COXO_BOVIN	P00430 bos taurus
3	21	100.0	63	1 COXO_MOUSE	P17665 mus musculus
4	21	100.0	90	1 MS2B_DROMA	P33738 drosophila
5	21	100.0	90	1 MS2B_DROME	P10334 drosophila
6	21	100.0	90	1 MS2B_DROSE	P33729 drosophila
7	21	100.0	90	1 MS2B_DROSI	P33740 drosophila
8	21	100.0	102	1 U185_ORYSA	Q94dm8 oryza sativ
9	21	100.0	102	1 YD86_MYCTU	P71656 mycobacteri
10	21	100.0	127	1 CAL2_HUMAN	P10092 homo sapien
11	21	100.0	127	1 Y355_TREPA	O83374 treponema p
12	21	100.0	128	1 CAL1_HUMAN	P06881 homo sapien
13	21	100.0	130	1 CAL2_MOUSE	O99mp3 mus musculus
14	21	100.0	130	1 SZ05_RAT	P37885 rattus norv
15	21	100.0	133	1 P2Y2_BOVIN	O18951 bos taurus
16	21	100.0	133	1 R8PA_SYNY3	P57885 synchocyst
17	21	100.0	134	1 CAL2_RAT	P10093 rattus norv
18	21	100.0	134	1 CUJA_TENMO	P80683 tenebrio mo
19	21	100.0	136	1 N2UM_ARTSA	P19042 artemia sal
20	21	100.0	141	1 CAL0_HUMAN	P01252 homo sapien
21	21	100.0	141	1 YEF5_YEAST	P32616 saccharomyc
22	21	100.0	142	1 PSAH_MAIZE	O65101 zea mays (m
23	21	100.0	143	1 PSAH_HOVRU	P20143 hordeum vul
24	21	100.0	144	1 PSAH_SPIOL	P22179 spinacia ol
25	21	100.0	145	1 PSAH_BRARA	O04006 brassica ra
26	21	100.0	145	1 PSH1_ARATH	Q9su17 arabidopsis
27	21	100.0	145	1 PSH1_ARATH	Q9su17 arabidopsis
28	21	100.0	165	1 VNS3_CVPPU	P09047 porcine tra
29	21	100.0	168	1 RBS_SACHY	Q41373 saccharum h
30	21	100.0	181	1 RBS_LACSA	O40250 lactuca sat
31	21	100.0	182	1 PAAD_ARCFU	O29054 archaeoglob
32	21	100.0	186	1 P152_METTH	O26255 methanobact
33	21	100.0	190	1 APT_TREPA	O84001 treponema p

34 21 100.0 195 1 VP12\_BPPH6 P07580 bacterioph  
35 21 100.0 200 1 ACD2\_CLOAB Q97dq1 clostridium  
36 21 100.0 200 1 RS8A\_SCHPO O14049 schizosacch  
37 21 100.0 200 1 RS8B\_SCHPO Q9p7b2 schizosacch  
38 21 100.0 202 1 YD99\_HAEIN P44175 haemophilus  
39 21 100.0 205 1 YWRF\_BACSU O05220 bacillus su  
40 21 100.0 206 1 YW80\_YEAST Q04019 saccharomyc  
41 21 100.0 207 1 GTA\_CAEEL Q09596 caenorhabdi  
42 21 100.0 208 1 RS8\_CAEEL P48156 caenorhabdi  
43 21 100.0 209 1 YW7B\_BACHD Q9kak9 bacillus ba  
44 21 100.0 210 1 RS6\_ARATH Q9fif3 arabidopsis  
45 21 100.0 217 1 YAEI\_ECOLI P31547 escherichia

## ALIGNMENTS

RESULT 1  
SUGA\_ACHDO STANDARD; PRT; 5 AA.  
AC P19991;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Suboesophageal ganglion pentapeptide.  
OS Acheta domesticus (House cricket).  
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Orthopteroidea; Orthoptera; Ensifera;  
OC Gryllidae; Gryllinae; Acheta.  
OX NCBI\_TaxID=6997;  
RN [1]  
RP SEQUENCE.  
RA Wicker C., Wicker C.:  
RT "Isolation and structure of a peptide isolated from the  
RT suboesophageal ganglion of Acheta domesticus (orthoptera).";  
RL Comp. Biochem. Physiol. 88C:185-187(1987).  
CC -!- SUBCELLULAR LOCATION: MAIN PEPTIDE FROM THE SUBESOPHAGEAL  
CC GANGLIA.  
DR PIR; JS0319; JS0319.  
SQ SEQUENCE 5 AA; 476 MW; 69D76DDDDDB00000 CRC64;

Query Match 100.0%; Score 21; DB 1; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.le+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
Db 2 AAPF 5

RESULT 2  
COXO\_BOVIN STANDARD; PRT; 63 AA.  
AC P00430;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cytochrome c oxidase polypeptide VIIC, mitochondrial precursor  
DE (EC 1.9.3.1) (VIIIA).  
GN COX7C OR COX7CP1.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Butheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90045968; PubMed=2554257;  
RA Aqua M.S., Lonax M.I., Schon E.A., Grossman L.I.;  
RT "Nucleotide sequence of a cDNA for bovine cytochrome c oxidase  
RT subunit VIIC.";  
RL Nucleic Acids Res. 17:8376-8376(1989).  
RN [2]

RP	SEQUENCE FROM N.A.	CC	EMBL; X15725; CAB57793.1; -
RC	TISSUE=Heart;	DR	EMBL; X58823; CAA41627.1; -
RA	Aqua M.S., Bachman N.J., Lomax M.I., Grosman L.I.;	DR	EMBL; U58655; AAC48719.1; -
RT	"Characterization and expression of a cDNA specifying subunit VIIC of	DR	PIR; JH0473; OSB08A.
RT	bovine cytochrome c oxidase."	DR	PDB; 10CC; 07-DEC-96.
RL	Gene 104:211-217(1991).	DR	PDB; 20CC; 13-JAN-99.
RN	[3]	DR	PDB; 10CR; 29-JUL-99.
RP	SEQUENCE FROM N.A.	DR	PDB; 10CO; 22-JUL-99.
RP	MEDLINE=97248578; PubMed=9092564;	DR	PDB; 10CZ; 22-JUL-99.
RA	Seelan R.S., Grosman L.I.;	DR	InterPro; IPR004202; COX7C.
RT	"Structural organization and promoter analysis of the bovine	KW	Pfam; PF02935; COX7C; 1.
RT	cytochrome c oxidase subunit VIIC gene. A functional role for YY1."	KW	Oxidoreductase; Mitochondrion; Transit peptide; 3D-structure;
RL	J. Biol. Chem. 272:10175-10181(1997).	KW	Inner membrane; Transmembrane.
RN	[4]	FT	TRANSIT 1 16
RP	SEQUENCE OF 17-63.	FT	CHAIN 17 63
RC	TISSUE=Heart;	FT	DOMAIN 17 33
RX	MEDLINE=79046803; PubMed=213363;	FT	TRANSMEM 34 60
RA	Buse G., Steffens G.J.;	FT	DOMAIN 61 63
RT	"Studies on cytochrome c oxidase, II. The chemical constitution of a	SQ	SEQUENCE 63 AA; 7331 MW; C6A2AD5CC4B63COB CRC64;
RT	short polypeptide from the beef heart enzyme."		
RL	Hoppe-Seyler's Z. Physiol. Chem. 359:1005-1009(1978).		
RN	[5]		
RP	SEQUENCE OF 17-42.		
RC	TISSUE=Liver;		
RX	MEDLINE=69000697; PubMed=2844245;		
RA	Yanamura W., Zhang Y.-Z., Takamiya S., Capaldi R.A.;		
RT	"Tissue-specific differences between heart and liver cytochrome c		
RT	oxidase."		
RL	Biochemistry 27:4909-4914(1988).		
RN	[6]		
RP	X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).		
RX	MEDLINE=96216288; PubMed=8638158;		
RA	Taukithara T., Aoyama H., Yamashita E., Tomizaki T., Yamaguchi H.,		
RA	Shinzawa-Itoh K., Nakashima R., Yaono R., Yoshikawa S.;		
RT	"The whole structure of the 13-subunit oxidized cytochrome c oxidase		
RT	at 2.8 A."		
RL	Science 272:1136-1144(1996).		
RN	[7]		
RP	X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).		
RC	TISSUE=Heart;		
RX	MEDLINE=99190827; PubMed=10089392;		
RA	Tomizaki T., Yamashita E., Yamaguchi H., Aoyama H., Tsukihara T.,		
RA	Shinzawa-Itoh K., Nakashima R., Yaono R., Yoshikawa S.;		
RT	"Structure analysis of bovine heart cytochrome c oxidase at 2.8 A		
RT	resolution."		
RL	Acta Crystallogr. D 55:31-45(1999).		
RN	[8]		
RP	X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).		
RC	TISSUE=Heart;		
RX	MEDLINE=20235645; PubMed=10771420;		
RA	Fei M.J., Yamashita E., Inoue N., Yao M., Yamaguchi H., Tsukihara T.,		
RA	Shinzawa-Itoh K., Nakashima R., Yoshikawa S.;		
RT	"X-ray structure of azide-bound fully oxidized cytochrome c oxidase		
RT	from bovine heart at 2.9 A resolution."		
RL	Acta Crystallogr. D 56:529-535(2000).		
CC	-1- FUNCTION: THIS PROTEIN IS ONE OF THE NUCLEAR-CODED POLYPEPTIDE		
CC	CHAINS OF CYTOCHROME C OXIDASE, THE TERMINAL OXIDASE IN		
CC	C + 2 H(2)O.		
CC	-1- CATALYTIC ACTIVITY: 4 ferrocycytochrome c + O(2) = 4 ferricytochrome		
CC	c + 2 H(2)O.		
CC	-1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE VIIC FAMILY.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	EMBL; X52940; CAA37115.1; -		
DR	PIR; S10303; S10303.		
DR	MGI; MGI103226; Cox7c.		
DR	InterPro; IPR004202; COX7C.		
DR	Pfam; PF02935; COX7C; 1.		
KW	Oxidoreductase; Mitochondrion; Transit peptide.		



```

FT TRANSIT 1 16 MITOCHONDRION.
FT CHAIN 17 63
SQ SEQUENCE 63 AA; 7333 MW; C6A2AD4BFEB2C7B CRC64;

Query Match 100.0%; Score 21; DB 1; Length 63;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4
Db 50 AAPF 53

RESULT 4
MS2B DROME STANDARD; PRT; 90 AA.
AC P33738
DT 01-FEB-1994 (Rel. 28, Created)
DE 01-FEB-1994 (Rel. 28, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Accessory gland-specific peptide 26Ab precursor (Male accessory
DE gland secretory protein 355B).
GN ACP26AB OR MST26AB OR MST355B.
OS Drosophila mauritiana (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7226;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Robertson C340;
RX MEDLINE=93106377; PubMed=1361475;
RA Aguade M., Miyashita N., Langley C.H.;
RT "Polymorphism and divergence in the Mst26A male accessory gland gene
RT region in Drosophila.";
RL Genetics 132:755-770(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Various strains;
RX MEDLINE=98384843; PubMed=9718731;
RA Tsaur S.-C., Ting C.-T., Wu C.-I.;
RT "Positive selection driving the evolution of a gene of male
RT reproduction, Acp26Aa, of Drosophila: II. Divergence versus
RT polymorphism.";
RL Mol. Biol. Evol. 15:1040-1046(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Various strains;
RX MEDLINE=98384843; PubMed=9718731;
RA Tsaur S.-C., Ting C.-T., Wu C.-I.;
RT "Positive selection driving the evolution of a gene of male
RT reproduction, Acp26Aa, of Drosophila: II. Divergence versus
RT polymorphism.";
RL Mol. Biol. Evol. 15:1040-1046(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=M036A, M037A, M047A, M079B, M080B, LA25, LA46, and LA108;
RX MEDLINE=99016087; PubMed=9799260;
RA Aguade M.;
RT "Differential forces drive the evolution of the Acp26Aa and Acp26Ab
RT accessory gland genes in the Drosophila melanogaster species
RT complex.";
RL Genetics 150:1079-1089(1998).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Achayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bensch P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jaisi M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

```

```

DE DE Accessory gland-specific peptide 26Ab precursor (Male accessory
DE gland secretory protein 355B).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=89053045; PubMed=3142802;
RA Monsma S.A., Wolfner M.F.;
RT "Structure and expression of a Drosophila male accessory gland gene
RT whose product resembles a peptide pheromone precursor.";
RL Genes Dev. 2:1063-1073(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NC1, NC3, NC4, NC5, NC6, NC7, NC8, NC9, and NC10;
RX MEDLINE=93106377; PubMed=1361475;
RA Aguade M., Miyashita N., Langley C.H.;
RT "Polymorphism and divergence in the Mst26A male accessory gland gene
RT region in Drosophila.";
RL Genetics 132:755-770(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Various strains;
RX MEDLINE=98384843; PubMed=9718731;
RA Tsaur S.-C., Ting C.-T., Wu C.-I.;
RT "Positive selection driving the evolution of a gene of male
RT reproduction, Acp26Aa, of Drosophila: II. Divergence versus
RT polymorphism.";
RL Mol. Biol. Evol. 15:1040-1046(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=M036A, M037A, M047A, M079B, M080B, LA25, LA46, and LA108;
RX MEDLINE=99016087; PubMed=9799260;
RA Aguade M.;
RT "Differential forces drive the evolution of the Acp26Aa and Acp26Ab
RT accessory gland genes in the Drosophila melanogaster species
RT complex.";
RL Genetics 150:1079-1089(1998).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Achayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bensch P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jaisi M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

```

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Scapleton M., Strong R., Sun E.,  
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
CC -1- FUNCTION: THIS PROTEIN IS TRANSFERRED FROM MALE TO FEMALE DURING  
CC MATING AND MAY AFFECT EGG-LAYING AND BEHAVIOR AFTER MATING.  
CC -1- SUBCELLULAR LOCATION: Extracellular.  
CC -1- TISSUE SPECIFICITY: MAIN CELLS OF THE ACCESSORY GLANDS OF MALES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; Y00219; CAA58367.1; -  
DR EMBL; X70888; CAA50233.1; -  
DR EMBL; X70889; CAA50235.1; -  
DR EMBL; X70890; CAA50237.1; -  
DR EMBL; X70891; CAA50239.1; -  
DR EMBL; X70892; CAA50241.1; -  
DR EMBL; X70893; CAA50243.1; -  
DR EMBL; X70894; CAA50245.1; -  
DR EMBL; X70895; CAA50247.1; -  
DR EMBL; X70896; CAA50249.1; -  
DR EMBL; X70897; CAA50251.1; -  
DR EMBL; AF052470; AAC27996.1; -  
DR EMBL; AF052471; AAC27998.1; -  
DR EMBL; AF052472; AAC28000.1; -  
DR EMBL; AF052473; AAC28002.1; -  
DR EMBL; AF052474; AAC28004.1; -  
DR EMBL; AF052475; AAC28006.1; -  
DR EMBL; AF052476; AAC28008.1; -  
DR EMBL; AF052477; AAC28010.1; -  
DR EMBL; AF052478; AAC28012.1; -  
DR EMBL; AF052479; AAC28014.1; -  
DR EMBL; AF052480; AAC28016.1; -  
DR EMBL; AF052481; AAC28018.1; -  
DR EMBL; AF053250; AAC28791.1; -  
DR EMBL; AF053251; AAC28793.1; -  
DR EMBL; AF053252; AAC28795.1; -  
DR EMBL; AF053253; AAC28797.1; -  
DR EMBL; AF053254; AAC28799.1; -  
DR EMBL; AF053255; AAC28801.1; -  
DR EMBL; AF053256; AAC28803.1; -  
DR EMBL; AF053257; AAC28805.1; -  
DR EMBL; AF053258; AAC28807.1; -  
DR EMBL; AF053259; AAC28809.1; -  
DR EMBL; AF053260; AAC28811.1; -  
DR EMBL; AF053261; AAC28813.1; -  
DR EMBL; AF053262; AAC28815.1; -  
DR EMBL; AF053263; AAC28817.1; -  
DR EMBL; AF053264; AAC28819.1; -  
DR EMBL; AF053265; AAC28821.1; -  
DR EMBL; AF053266; AAC28823.1; -  
DR EMBL; AF053267; AAC28825.1; -  
DR EMBL; AF053268; AAC28827.1; -  
DR EMBL; AF053269; AAC28829.1; -  
DR EMBL; AF053270; AAC28831.1; -  
DR EMBL; AF053271; AAC28833.1; -  
DR EMBL; AF053272; AAC28835.1; -  
DR EMBL; AF053273; AAC28837.1; -  
DR EMBL; AF053274; AAC28839.1; -  
DR EMBL; AF053275; AAC28841.1; -

DR EMBL; AF053276; AAC28843.1; -  
DR EMBL; AJ231353; CAB37201.1; -  
DR EMBL; AJ231354; CAB37203.1; -  
DR EMBL; AJ231356; CAB37207.1; -  
DR EMBL; AJ231358; CAB37211.1; -  
DR EMBL; AJ231359; CAB37213.1; -  
DR EMBL; AJ231365; CAB37225.1; -  
DR EMBL; AJ231372; CAB37239.1; -  
DR EMBL; AJ231379; CAB37251.1; -  
DR EMBL; AE003611; AAF52297.1; -  
DR PIR; S02854; S02854.  
DR PIR; S30408; S30408.  
DR PIR; S30410; S30410.  
DR PIR; S30424; S30424.  
DR FlyBase; FBgn0002856; Acp26Ab.  
KW Behavior; Signal; Polymorphism.  
FT SIGNAL 1 21  
FT CHAIN 22 90  
FT VARIANT 64 64  
FT  
FT VARIANT 77 77  
FT VARIANT 80 80  
FT  
FT VARIANT 89 89  
FT VARIANT 90 90  
FT  
FT SEQUENCE 90 AA; 10162 MW; 1FDFDC7951F9BDF6 CRC64;  
SQ  
Query Match 100.0%; Score 21; DB 1; Length 90;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 AAPF 4  
Db 21 AAPF 24  
  
RESULT 6  
MS2B\_DROSE STANDARD; PRT; 90 AA.  
AC F33739;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Accessory gland-specific peptide 26Ab precursor (Male accessory  
DE gland secretory protein 355B).  
GN ACP26AB OR MST26AB OR MST355B.  
OS *Drosophila sechellia* (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7238;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93106377; PubMed=1361475;  
RA Aguade M., Miyashita N., Langley C.H.;  
RT "Polymorphism and divergence in the Mat26A male accessory gland gene  
RL region in *Drosophila*.";  
CC Genetics 132:755-770(1992).  
CC -1- FUNCTION: THIS PROTEIN IS TRANSFERRED FROM MALE TO FEMALE DURING  
CC MATING AND MAY AFFECT EGG-LAYING AND BEHAVIOR AFTER MATING.  
CC -1- SUBCELLULAR LOCATION: Extracellular.  
CC -1- TISSUE SPECIFICITY: MAIN CELLS OF THE ACCESSORY GLANDS OF MALES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X72630; CAA51207.1; -  
 DR FlyBase; FBgn0012778; Dsec\Acp26Ab.  
 KW Behavior; Signal.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 90 ACCESSORY GLAND-SPECIFIC PEPTIDE 26AB.  
 SQ SEQUENCE 90 AA; 10270 MW; 99FC4EE69B7782E CRC64;  
 Query Match 100.0%; Score 21; DB 1; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AAPF 4  
 Db |||||  
 Db 21 AAPF 24  
 RESULT 7  
 MS2B\_DROSI  
 ID MS2B\_DROSI STANDARD; PRT; 90 AA.  
 AC P33740;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Accessory gland-specific peptide 26Ab precursor (Male accessory  
 DE Gland secretory protein 355B).  
 GN ACP26AB OR MST26AB OR MST355B.  
 OS Drosophila simulans (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7240;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PC;  
 RA Aguade M., Miyashita N., Langley C.H.;  
 RT "Polymorphism and divergence in the Mst26A male accessory gland gene  
 RT region in Drosophila.";  
 RL Genetics 132:755-770(1992).  
 CC -!- FUNCTION: THIS PROTEIN IS TRANSFERRED FROM MALE TO FEMALE DURING  
 CC MATING AND MAY AFFECT EGG-LAYING AND BEHAVIOR AFTER MATING.  
 CC -!- SUBCELLULAR LOCATION: Extracellular.  
 CC -!- TISSUE SPECIFICITY: MAIN CELLS OF THE ACCESSORY GLANDS OF MALES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X70899; CAA50255.1; -  
 DR PIR; S30430; S30430.  
 DR FlyBase; FBgn0012821; Dslm\Acp26Ab.  
 KW Behavior; Signal.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 90 ACCESSORY GLAND-SPECIFIC PEPTIDE 26AB.  
 SQ SEQUENCE 90 AA; 10189 MW; F7F8D6E6D04F03E5 CRC64;  
 Query Match 100.0%; Score 21; DB 1; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AAPF 4  
 Db |||||  
 Db 21 AAPF 24  
 RESULT 8  
 PIR

U185\_ORYSA  
 ID U185\_ORYSA STANDARD; PRT; 102 AA.  
 AC Q94DM8;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein P0483G10.15.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 RT clone:P0483G10.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO THE UPF0185 (BM-002) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF003263; BAB63621.1; -  
 DR InterPro; IPR005375; UPF0185.  
 DR Pfam; PF03671; UPF0185; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 102 AA; 10362 MW; 1E8F85B27F371159 CRC64;  
 Query Match 100.0%; Score 21; DB 1; Length 102;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AAPF 4  
 Db |||||  
 Db 99 AAPF 102  
 RESULT 9  
 YD86\_MYCTU  
 ID YD86\_MYCTU STANDARD; PRT; 102 AA.  
 AC P71656;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein Rv1386 precursor.  
 GN Rv1386 OR MT1430 OR MTCV21B4.03.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RA MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 RN [2]

```

RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Allard D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Dalcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Blehai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z80108; CAB02191.1; -.
CC EMBL; AE007014; AAK45695.1; -.
CC TIGR; MT1430; -.
CC TubercuList; Rv1386; -.
CC InterPro; IPR000084; PE_1; Signal; Complete proteome.
CC Pfam; PF00934; PE; 1; Signal; Complete proteome.
CC Hypothetical protein; Signal; Complete proteome.
CC SIGNAL 1 31 POTENTIAL.
CC CHAIN 32 102 HYPOTHETICAL PROTEIN RV1386.
CC SEQUENCE 102 AA; 9862 MW; 8165F09D8DB9D752 CRC64;
CC -----
Query Match 100.0%; Score 21; DB 1; Length 102;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 AAPF 4
Db 31 AAPF 34
RESULT 10
ID CAL2 HUMAN STANDARD; PRT; 127 AA.
AC P10092; Q9UCN9;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Calcitonin gene-related peptide II precursor (CGRP-II) (Beta-type
DE CGRP).
GN CALCB OR CALC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8710523; PubMed=3492393;
RA Steenbergh P.H., Hoepfner J.W.M., Zandberg J., Visser A.,
RA Lips C.J.M., Jansz H.S.;
RT "Structure and expression of the human calcitonin/CGRP genes.";
RL FEBS Lett. 209:97-103(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=8710523; PubMed=3492393;
RA Steenbergh P.H., Hoepfner J.W.M., Zandberg J., Visser A.,
RA Lips C.J.M., Jansz H.S.;
RT "Structure and expression of the human calcitonin/CGRP genes.";
RL FEBS Lett. 209:97-103(1986).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=8710523; PubMed=3492393;
RA Steenbergh P.H., Hoepfner J.W.M., Zandberg J., Lips C.J.M.,
RA Jansz H.S.;
RT "A second human calcitonin/CGRP gene.";
RL FEBS Lett. 183:403-407(1985).
RN [4]
RP PARTIAL SEQUENCE OF 82-108.
RX MEDLINE=87109142; PubMed=3492492;
RA Petermann J.B., Born W., Chang J.Y., Fischer J.A.;
RT "Identification in the human central nervous system, pituitary, and
RT thyroid of a novel calcitonin gene-related peptide, and partial amino
RT acid sequence in the spinal cord.";
RL J. Biol. Chem. 262:542-545(1987).
RN [5]
RP SEQUENCE OF 82-86 AND 104-117.
RX MEDLINE=90211348; PubMed=2322289;
RA Wimalawansa S.J., Morris H.R., Etienne A., Blench I., Panico M.,
RA McIntyre I.;
RT "Isolation, purification and characterization of beta-hCGRP from
RT human spinal cord.";
RL Biochem. Biophys. Res. Commun. 167:993-1000(1990).
RN [6]
RP SEQUENCE OF 82-104.
RX MEDLINE=92287083; PubMed=1318039;
RA Kitamura K., Kangawa K., Kawamoto M., Ichiki Y., Matsuo H., Eto T.;
RT "Isolation and characterization of peptides which act on rat
RT platelets, from a pheochromocytoma.";
RL Biochem. Biophys. Res. Commun. 185:134-141(1992).
CC -1- FUNCTION: CGRP induces vasodilatation. It dilates a variety of
CC vessels including the coronary, cerebral and systemic vasculature.
CC Its abundance in the CNS also points toward a neurotransmitter or
CC neuromodulator role.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE CALCITONIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X04855; CAC05295.1; -.
CC EMBL; X04857; CAC05295.1; JOINED.
CC EMBL; X04861; CAC05295.1; JOINED.
CC EMBL; BC008428; AAH08428.1; -.
CC PIR; X02404; CAA32429.1; -.
CC PIR; A25864; A25864.
CC PIR; A34565; A34565.
CC PIR; B26142; B26142.
CC Genew; HGNC:1438; CALCB.
CC MIM; 114160; -.
CC InterPro; IPR001693; Calcitonin-like.
CC InterPro; IPR002163; Calcitonin B.
CC Pfam; PF00214; Calc CGRP IAPP; 1.
CC PRINTS; PR00817; CALCITONINB.
CC SMART; SM00113; CALCITONIN; 1.
CC PROSITE; PS00258; CALCITONIN; 1.
CC Cleavage on pair of basic residues; Amidation; Hormone; Signal.
CC SIGNAL 1 25 POTENTIAL.
CC PROPEP 26 79 CALCITONIN GENE-RELATED PEPTIDE II.
CC PEPTIDE 82 118
CC PROPEP 124 127
CC DISULFID 83 88
CC MOD RES 118 118 AMIDATION (G-119 PROVIDE AMIDE GROUP).
CC CONFLICT 73 73 G -> S (IN REF. 3).
CC SEQUENCE 127 AA; 13706 MW; B0A71A063CD5ACE7 CRC64;
Query Match 100.0%; Score 21; DB 1; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 AAPF 4
Db 25 AAPF 28

```

```

RESULT 11
Y355 TREPA
ID Y355 TREPA STANDARD; PRT; 127 AA.
AC O83174
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein TP0355.
GN TP0355.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uterback T.,
RA McDonald L., Riattach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete."
RL Science 281:375-388(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC -----
DR EMBL; AE001215; AAC65353.1; --
DR TIGR; TP0355; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 127 AA; 14526 MW; 1B848EF0AC4ADC4 CRC64;
Query Match 100.0%; Score 21; DB 1; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAFP 4
Db 85 AAFP 88
RESULT 12
CALI HUMAN
ID CALI HUMAN STANDARD; PRT; 128 AA.
AC P06881; Q93048; Q9UCP0;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Calcitonin gene-related peptide 1 precursor (CGRP-I) (Alpha-type
DE CGRP).
GN CALCA OR CALCI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89386053; PubMed=25711128;
RA Broad P.M., Symes A.J., Thakker R.V., Craig R.K.;
RT "Structure and methylation of the human calcitonin/alpha-CGRP gene.";
RL Nucleic Acids Res. 17:6999-7011(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85166259; PubMed=3872459;
RA Jonas V., Lin C.R., Kawashima E., Semon D., Swanson L.W.,
RA Mermod J.-J., Evans R.M., Rosenfeld M.G.;
RT "Alternative RNA processing events in human calcitonin/calcitonin
RT gene-related peptide gene expression.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:1994-1998(1985).
RN [3]
RP SEQUENCE OF 48-119 FROM N.A.
RX MEDLINE=85022523; PubMed=6148938;
RA Nelkin B.D., Rosenfeld K.I., de Bustros A., Leong S.S., Roos B.A.,
RA Baylin S.B.;
RT "Structure and expression of a gene encoding human calcitonin and
RT calcitonin gene related peptide.";
RL Biochem. Biophys. Res. Commun. 123:648-655(1984).
RN [4]
RP SEQUENCE OF 49-128 FROM N.A.
RX MEDLINE=85230541; PubMed=2408883;
RA Edbrooke M.R., Parker D., McVey J.H., Riley J.H., Sorenson G.D.,
RA Pattengill O.S., Craig R.K.;
RT "Expression of the human calcitonin/CGRP gene in lung and thyroid
RT carcinoma.";
RL EMBO J. 4:715-724(1985).
RN [5]
RP SEQUENCE OF 77-128 FROM N.A.
RC TISSUE=Thyroid carcinoma;
RX MEDLINE=84240176; PubMed=6610687;
RA Steenbergh P.H., Hoppener J.W., Zandberg J., de Ven W.J., Jansz H.S.,
RA Lips C.J.;
RT "Calcitonin gene related peptide coding sequence is conserved in the
RT human genome and is expressed in medullary thyroid carcinoma.";
RL J. Clin. Endocrinol. Metab. 59:358-360(1984).
RN [6]
RP SEQUENCE OF 77-128 FROM N.A.
RX MEDLINE=87213363; PubMed=3034287;
RA Craig R.K., Riley J.H., Edbrooke M.R., Broad P.M., Poord S.M.,
RA Al-Kazwini S.J., Holman J.J., Marshall I.;
RT "Expression and function of the human calcitonin/alpha-CGRP gene in
RT health and disease.";
RL Biochem. Soc. Symp. 52:91-105(1986).
RN [7]
RP SEQUENCE OF 83-119.
RX MEDLINE=84191466; PubMed=6609312;
RA Morris H.R., Panico M., Etienne T., Tippins J., Girgis S.I.,
RA McIntyre I.;
RT "Isolation and characterization of human calcitonin gene-related
RT peptide.";
RL Nature 308:746-748(1984).
RN [8]
RP PARTIAL SEQUENCE OF 83-117.
RX MEDLINE=87109142; PubMed=3492492;
RA Petermann J.B., Born W., Chang J.Y., Fischer J.A.;
RT "Identification in the human central nervous system, pituitary, and
RT thyroid of a novel calcitonin gene-related peptide, and partial amino
RT acid sequence in the spinal cord.";
RL J. Biol. Chem. 262:542-545(1987).
RN [9]
RP SEQUENCE OF 83-108, AND FUNCTION.
RC TISSUE=Pheochromocytoma;
RX MEDLINE=92287083; PubMed=1318039;
RA Kitamura K., Kangawa K., Kawamoto M., Ichiki Y., Matsuo H., Eto T.;
RT "Isolation and characterization of peptides which act on rat
RT platelets, from a pheochromocytoma.";
RL Biochem. Biophys. Res. Commun. 185:134-141(1992).
RN [10]
RP STRUCTURE BY NMR OF CGRP.
RX MEDLINE=91105142; PubMed=1988044;
RA Breeze A.L., Harvey T.S., Bazzo R., Campbell I.D.;
RT "Solution structure of human calcitonin gene-related peptide by 1H
RT NMR and distance geometry with restrained molecular dynamics.";
RL Biochemistry 30:575-582(1991).
RN [11]
RP STRUCTURE BY NMR OF CGRP.
RX MEDLINE=91248117; PubMed=2039456;

```

```

RA Hubbard J.A.M., Martin S.R., Chaplin L.C., Bose C., Kelly S.M.,
RA Price N.C.;
RA "Solution structures of calcitonin-gene-related-peptide analogues of
RT calcitonin-gene-related peptide and amylin.";
RL Biochem. J. 275:785-788(1991).
CC
CC -1- FUNCTION: CGRP induces vasodilatation. It dilates a variety of
CC vessels including the coronary, cerebral and systemic vasculature.
CC Its abundance in the CNS also points toward a neurotransmitter or
CC neuromodulator role. It also elevates platelet CAMP.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (AC P01258), 2 (AC P01258) AND
CC 3 (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE CALCITONIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X15943; CAA34070.1; -
CC EMBL; X02330; CAA26190.1; -
CC EMBL; M12667; AAA51914.1; -
CC EMBL; M12667; AAA51914.1; JOINED.
CC EMBL; M12665; AAA51914.1; JOINED.
CC EMBL; K03512; AAA52011.1; -
CC EMBL; M28637; AAA52012.1; -
CC EMBL; M26094; AAA51912.1; -
CC PIR; B22716; TCHUR.
CC PIR; A22949; A22949.
CC PIR; A26142; A26142.
CC PIR; S07644; S07644.
CC Genew; HGNC:1437; CALCA.
CC MIM; 114130; -
CC InterPro; IPR001693; Calcitonin-like.
CC InterPro; IPR002163; Calcitonin B.
CC Pfam; PF00214; Calc CGRP IAPP; I.
CC PRINTS; PR00817; CALCITONINB.
CC SMART; SM00113; CALCITONIN; 1.
CC PROSITE; PS00258; CALCITONIN; 1.
CC Cleavage on pair of basic residues; Amidation; Alternative splicing;
CC Hormone; Signal.
CC SIGNAL 1 25 POTENTIAL.
CC PROPEP 26 80
CC PEPTIDE 83 119 CALCITONIN GENE-RELATED PEPTIDE I.
CC PROPEP 125 128
CC DISULFID 84 89
CC MOD_RES 119 119
CC CONFLICT 48 48
CC CONFLICT 76 76 R -> V (IN REF. 3).
CC CONFLICT 76 76 R -> V (IN REF. 4).
CC SEQUENCE 128 AA; 13999 MW; A003A1069260D9B8 CRC64;
SQ
Query Match 100.0%; Score 21; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAPF 4
DB 25 AAPF 28
RESULT 13
CAL2 MOUSE
ID CAL2_MOUSE STANDARD; PRT; 130 AA.
AC Q99MP3;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Calcitonin gene-related peptide II precursor (CGRP-II) (Beta-type
DE CGRP).
OS CALCB.
OS Mus musculus (Mouse).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RC MEDLINE=21604266; PubMed=11761712;
RA Thomas P.M., Nasonkin I., Zhang H., Gagel R.F., Cote G.J.;
RT "Structure of the mouse calcitonin/calcitonin gene-related peptide
RT alpha and beta genes.";
RL DNA Seq. 12:131-135(2001).
CC -1- FUNCTION: CGRP induces vasodilatation. It dilates a variety of
CC vessels including the coronary, cerebral and systemic vasculature.
CC Its abundance in the CNS also points toward a neurotransmitter or
CC neuromodulator role (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE CALCITONIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF325526; AAK16431.1; -
CC EMBL; AF325524; AAK16431.1; JOINED.
CC MGD; MGI:2151254; Calcb.
CC InterPro; IPR000443; Amylin.
CC InterPro; IPR001693; Calcitonin-like.
CC InterPro; IPR002163; Calcitonin B.
CC Pfam; PF00214; Calc CGRP IAPP; I.
CC PRINTS; PR00817; CALCITONIN.
CC PRINTS; PR00818; ISLETAMYL0ID.
CC SMART; SM00113; CALCITONIN; 1.
CC PROSITE; PS00258; CALCITONIN; 1.
CC Cleavage on pair of basic residues; Amidation; Hormone; Signal.
CC SIGNAL 1 26 POTENTIAL.
CC PROPEP 27 82 BY SIMILARITY.
CC PEPTIDE 84 120 CALCITONIN GENE-RELATED PEPTIDE II.
CC PROPEP 127 130 BY SIMILARITY.
CC DISULFID 85 90 BY SIMILARITY.
CC MOD_RES 120 120
CC SEQUENCE 130 AA; 14623 MW; 97299244E8F6C536 CRC64;
SQ
Query Match 100.0%; Score 21; DB 1; Length 130;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAPF 4
DB 26 AAPF 29
RESULT 14
SZ05 RAT
ID SZ05_RAT STANDARD; PRT; 130 AA.
AC P97885;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DB Small inducible cytokine B5 precursor (CXCL5) (Cytokine LIX).
GN SCVB5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RC Kelner G.S., Maciejewski-Lenoir D., Lee E.D., Maki R.A.;
RA

```

```

RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY PARTICIPATE IN THE RECRUITMENT OF INFLAMMATORY CELLS
CC -!- BY INJURED OR INFECTED TISSUE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INTERCRINE ALPHA FAMILY (SMALL CYTOKINE
CC C-X-C) (CHEMOKINE CXCL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U90448; AAB61460.1; -.
DR HSSP; P10889; IMI2.
DR InterPro; IPR001089; CXC_chemokine_gml1.
DR InterPro; IPR001811; Chemokine_IL8.
DR Pfam; PF00048; IL8; 1.
DR PRINTS; PR00437; SMALLCYTCKXC.
DR SMART; SM00199; SCY; 1.
DR PROSITE; PS00471; SMALL_CYTOKINES_CXC; 1.
KW Cytokine; Signal.
FT SIGNAL 1 37 POTENTIAL.
FT CHAIN 38 130 SMALL INDUCIBLE CYTOKINE B5.
FT DISULFID 50 76 BY SIMILARITY.
FT DISULFID 52 93 BY SIMILARITY.
SQ SEQUENCE 130 AA; 14263 MW; C00F6B3605524F4E CRC64;
Query Match 100.0%; Score 21; DB 1; Length 130;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAPP 4
DB 37 AAPF 40
RESULT 15
P2Y2_BOVIN
ID P2Y2_BOVIN STANDARD; PRT; 133 AA.
AC O18951;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE P2Y purinoceptor 2 (P2Y2) (Fragment).
GN P2Y2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Chang A.S., Chang S.M., Schilling W.P.;
RT "Identification of multiple P2 purinergic receptor subtypes in bovine
RT aortic endothelial cells (BAECs).";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Receptor for ATP and UTP coupled to G-proteins that
CC activate a phosphatidylinositol-calcium second messenger system
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF005153; AAB62543.1; -.

```

```

DR InterPro; IPR000276; GPCR_Rhodopsin.
DR Pfam; PF00001; 7tm_1; 1.
DR PROSITE; PS00237; G_PROTEIN_RECP_F1_1; FALSE_NEG.
DR PROSITE; PS0262; G_PROTEIN_RECP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane.
FT NON_TER 1 1
FT DOMAIN 1 26 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 27 47 4 (POTENTIAL).
FT DOMAIN 48 74 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 75 95 5 (POTENTIAL).
FT DOMAIN 96 116 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 117 >133 6 (POTENTIAL).
FT NON_TER 133 133
SQ SEQUENCE 133 AA; 15006 MW; 301B087C62C75B69 CRC64;
Query Match 100.0%; Score 21; DB 1; Length 133;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAPP 4
DB 82 AAPF 85

```

Search completed: February 12, 2003, 10:23:20  
Job time : 2.53731 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:16:06 ; Search time 2.82985 Seconds  
(without alignments)  
291.248 Million cell updates/sec

Title: US-10-036-371-7

Perfect score: 21

Sequence: 1 AAPF 4

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mmc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp Vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	ID	Description
1	21	100.0	41 10 Q9S9F1	Q9S9F1 brassica na
2	21	100.0	45 2 P70957	P70957 bacillus st
3	21	100.0	47 6 Q9TRE9	Q9TRE9 sus scrofa
4	21	100.0	47 6 Q9TQX9	Q9TQX9 bos taurus
5	21	100.0	50 16 Q8YCK3	Q8YCK3 bruceella me
6	21	100.0	51 4 Q9Y5N3	Q9Y5N3 homo sapien
7	21	100.0	57 10 Q9W7N1	Q9W7N1 triticum ae
8	21	100.0	59 16 Q9ADG2	Q9ADG2 streptomyc
9	21	100.0	69 10 Q8S4B7	Q8S4B7 amblystegiu
10	21	100.0	70 12 Q9DWS4	Q9DWS4 hepatitis c
11	21	100.0	72 7 Q9MWJ2	Q9MWJ2 certhidea o
12	21	100.0	73 5 Q26512	Q26512 schistosoma
13	21	100.0	75 16 Q8UEW5	Q8UEW5 agrobacteri
14	21	100.0	76 13 Q9DFN9	Q9DFN9 gilllichthys
15	21	100.0	77 16 Q98L67	Q98L67 rhizobium l
16	21	100.0	78 10 Q8S493	Q8S493 amblystegiu

17	21	100.0	79 10 Q8S4B6	Q8S4B6 amblystegiu
18	21	100.0	79 10 Q8S4B4	Q8S4B4 amblystegiu
19	21	100.0	79 10 Q8S4B2	Q8S4B2 amblystegiu
20	21	100.0	79 10 Q8S4A7	Q8S4A7 amblystegiu
21	21	100.0	79 10 Q8S499	Q8S499 amblystegiu
22	21	100.0	79 17 Q26308	Q26308 methanobact
23	21	100.0	80 16 Q97LM6	Q97LM6 clostridium
24	21	100.0	81 2 Q51347	Q51347 pseudomonas
25	21	100.0	81 7 P79449	P79449 bos taurus
26	21	100.0	81 7 P79450	P79450 bos taurus
27	21	100.0	83 12 Q9DKP9	Q9DKP9 hepatitis b
28	21	100.0	86 16 Q8X3A0	Q8X3A0 escherichia
29	21	100.0	87 10 Q9FSP9	Q9FSP9 oryza sativ
30	21	100.0	88 16 Q8UHU9	Q8UHU9 agrobacteri
31	21	100.0	90 2 Q923D9	Q923D9 escherichia
32	21	100.0	90 5 Q95NY3	Q95NY3 drosophila
33	21	100.0	90 5 Q95NY2	Q95NY2 drosophila
34	21	100.0	90 5 Q95NY1	Q95NY1 drosophila
35	21	100.0	90 5 Q97184	Q97184 drosophila
36	21	100.0	90 5 Q97185	Q97185 drosophila
37	21	100.0	90 5 Q97456	Q97456 drosophila
38	21	100.0	90 5 Q97460	Q97460 drosophila
39	21	100.0	92 2 Q9ANK5	Q9ANK5 bradyrhizob
40	21	100.0	92 7 Q95470	Q95470 bos indicus
41	21	100.0	94 16 Q8U4Y1	Q8U4Y1 agrobacteri
42	21	100.0	96 10 Q8S495	Q8S495 amblystegiu
43	21	100.0	97 10 Q8S4B1	Q8S4B1 amblystegiu
44	21	100.0	97 11 Q9CVJ8	Q9CVJ8 mus musculu
45	21	100.0	97 16 Q92EA6	Q92EA6 listeria in

#### ALIGNMENTS

#### RESULT 1

Q9S9F1 ID Q9S9F1 PRELIMINARY; PRT; 41 AA.  
AC Q9S9F1;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE Napsin short chain S4=CALMODULIN antagonist/calcium-dependent protein  
DE kinase substrate.  
OS Brassica napus (Rape).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Brassica.  
OX NCBI\_TaxID=3708;  
RN [1]\_TaxID=3708;  
RP SEQUENCE.  
RX MEDLINE=96283790; PubMed=8679670;  
RA Neumann G.M., Condon R., Thomas I., Polya G.M.;  
RT "Purification and sequencing of multiple forms of Brassica napus seed  
RT napsin small chains that are calmodulin antagonists and substrates for  
RT plant calcium-dependent protein kinase.",  
RL Biochim. Biophys. Acta 1295:23-33(1996);  
SQ SEQUENCE 41 AA; 4652 MW; 8BC6738503380553 CRC64;

Query Match 100.0%; Score 21; DB 10; Length 41;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
|||  
DB 2 AAPF 5

#### RESULT 2

P70957 ID P70957 PRELIMINARY; PRT; 45 AA.  
AC P70957;  
DT 01-FEB-1997 (TrEMBLrel. 02, Created)  
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)

DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
 DE Alpha-amylase precursor (Fragment).  
 GN AMYS  
 OS Bacillus stearothermophilus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Geobacillus.  
 OX NCBI\_TaxID=1422;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9102499; PubMed=2265757;  
 RA Jorgensen P.L., Hansen C.K., Poulsen G.B., Diderichsen B.;  
 RT "In vivo genetic engineering: homologous recombination as a tool for  
 RL plasmid construction.";  
 RL Gene 96:37-41(1990).  
 DR ENBL; M62638; AAA22242.1; -.  
 KW Signal.  
 FT SIGNAL 1 34 POTENTIAL.  
 FT CHAIN 35 >45 POTENTIAL.  
 FT NON TER 45 45  
 SQ SEQUENCE 45 AA; 5119 MW; 192F623EA13E6DA6 CRC64;  
  
 Query Match 100.0%; Score 21; DB 2; Length 45;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 AAPF 4  
 Db 35 AAPF 38  
  
 RESULT 3  
 Q9TR9 PRELIMINARY; PRT; 47 AA.  
 ID Q9TR9  
 AC Q9TR9  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Cytochrome C oxidase subunit VIIC (Fragment).  
 OS Sus scrofa (pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=93384597; PubMed=8396926;  
 RA Sillard R., Jornvall H., Mutt V.;  
 RT "Characterization of porcine intestinal cytochrome c oxidase subunit  
 RL VIIC, purified by affinity chromatography.";  
 RL Biochem. Biophys. Res. Commun. 195:746-750(1993).  
 DR InterPro; IPR004202; COX7C.  
 DR Pfam; PF02935; COX7C; 1.  
 SQ SEQUENCE 47 AA; 5457 MW; E11E89CFBC18B5D7 CRC64;  
  
 Query Match 100.0%; Score 21; DB 6; Length 47;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 AAPF 4  
 Db 34 AAPF 37  
  
 RESULT 4  
 Q9TOX9 PRELIMINARY; PRT; 47 AA.  
 ID Q9TOX9  
 AC Q9TOX9  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Deoxyribonuclease P (Fragment).  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=94183469; PubMed=8136077;  
 RA Chang Y.M., Lin S., Liao T.H.;  
 RT "Bovine pancreatic deoxyribonuclease P: isoelectric focusing, peptide  
 RL mapping and primary structure.";  
 RL Biotechnol. Appl. Biochem. 19:129-140(1994).  
 DR HSSP; P00639; 3DNI  
 DR InterPro; IPR001582; DNase\_I.  
 DR PRINTS; PR00130; DNASEI.  
 DR ProDom; PD005408; DNaseI; 1.  
 SQ SEQUENCE 47 AA; 4937 MW; D757A9F441095799 CRC64;  
  
 Query Match 100.0%; Score 21; DB 6; Length 47;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 AAPF 4  
 Db 17 AAPF 20  
  
 RESULT 5  
 Q8YCK3 PRELIMINARY; PRT; 50 AA.  
 ID Q8YCK3  
 AC Q8YCK3  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical protein BMEI10525.  
 GN BMEI10525.  
 OS Brucella melitensis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Brucellaceae; Brucella.  
 OX NCBI\_TaxID=29459;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=16M / ATCC 23456 / BIOTYPE 1;  
 RX MEDLINE=20020109; PubMed=1175688;  
 RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mijer C., Los T.,  
 RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,  
 RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,  
 RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,  
 RA Haselkorn R., Kyrides N., Overbeek R.;  
 RT "The genome sequence of the facultative intracellular pathogen  
 RL Brucella melitensis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).  
 DR EMBL; AE009688; AAL53767.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 50 AA; 5208 MW; 8DE59A82243EFC17 CRC64;  
  
 Query Match 100.0%; Score 21; DB 16; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 AAPF 4  
 Db 44 AAPF 47  
  
 RESULT 6  
 Q9Y5N3 PRELIMINARY; PRT; 51 AA.  
 ID Q9Y5N3  
 AC Q9Y5N3  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Lecithin-cholesterol acyltransferase (EC 2.3.1.43) (Fragment).  
 GN LCAT.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia: Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ou J.F., Saku K., Liao Y.L., Jimi S., Jin W.J., Arakawa K.;
RT "A novel missense variant Thr13Met in the lecithin: cholesterol
RT acyltransferase (LCAT) gene associated with fish eye disease.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF140503; AAD28484.1; -
KW Acyltransferase; Transferase.
FT VARIANT 37 37 M -> T.
FT NON_TER 51 51
SQ SEQUENCE 51 AA; 5666 MW; EB511439A5129965 CRC64;

Query Match 100.0%; Score 21; DB 4; Length 51;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4
Db 22 AAPF 25

RESULT 7
Q9M7N1 PRELIMINARY; PRT; 57 AA.
AC Q9M7N1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE MFPI attachment factor 1 (Fragment).
GN MFPI.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99418887; PubMed=10489241;
RA Gundullis F., Peffer N.J., Meier I.;
RT "WAF1, a novel plant protein interacting with matrix attachment region
RT binding protein MFPI, is located at the nuclear envelope.";
RL Plant Cell 11:1755-1768(1999).
DR EMBL; AF181118; AAF6362.1; -
FT NON_TER 57 57
SQ SEQUENCE 57 AA; 6085 MW; E35F7AC99E250F78 CRC64;

Query Match 100.0%; Score 21; DB 10; Length 57;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4
Db 17 AAPF 20

RESULT 8
Q9ADG2 PRELIMINARY; PRT; 59 AA.
AC Q9ADG2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Hypothetical protein SCO6101.
GN SCO6101 OR SCBAC1A6.25C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;

Mammalia: Eutheria; Primates; Catarrhini; Hominidae; Homo.
Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL589708; CAC33946.1; -
KW Hypothetical protein.
SQ SEQUENCE 59 AA; 6387 MW; 885755E65DB6C9EF CRC64;

Query Match 100.0%; Score 21; DB 16; Length 59;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4
Db 5 AAPF 8

RESULT 9
Q8S4B7 PRELIMINARY; PRT; 69 AA.
AC Q8S4B7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Adenosine kinase (Fragment).
OS Amblystegium humile
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
OC Bryopsida; Bryidae; Hypnales; Amblystegiaceae; Amblystegium.
OX NCBI_TaxID=111433;
RN [1]
RP SEQUENCE FROM N.A.
RA Vanderpoorten A., Shaw A.J., Cox C.J.;
RT "Reconciled gene trees, molecular evolution, and species relationships
RT in Amblystegium.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF465020; AAL96441.1; -
KW Kinase.
FT NON_TER 1 1
FT NON_TER 69 69
SQ SEQUENCE 69 AA; 7793 MW; 81C1C22B4EAC30AC CRC64;

Query Match 100.0%; Score 21; DB 10; Length 69;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4
Db 40 AAPF 43

RESULT 10
Q9DWS4 PRELIMINARY; PRT; 70 AA.
AC Q9DWS4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE Hypothetical protein [Contains: envelope glycoprotein E1 (GP32) (GP35)]
DE (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=VE3216;
RX MEDLINE=20394384; PubMed=10935993;
RA Argentin C., Dettori S., Villano U., Guadagnino V., Infantolino D.,
RA Dantico P., Coppola R.C., Rapicetta M.;
RT "Molecular characterization of HCV genotype 4 isolates circulating in
  Italy.";
RL J. Med. Virol. 62:94-90(2000).
CC -1- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; AJ250216; CAC16983.1; -.
DR InterPro; IPR002519; HCV Env.
DR Pfam; PFO1539; HCV env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
  Transmembrane.
FT NON_TER 70 1
FT NON_TER 70 70
SQ SEQUENCE 70 AA; 7698 MW; BD360B781C3BF80F CRC64;

Query Match 100.0%; Score 21; DB 12; Length 70;
Best Local Similarity 100.0%; Pred. No. 5.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4
Db 54 AAPF 57

RESULT 11
Q9MWJ2 PRELIMINARY; PRT; 72 AA.
ID Q9MWJ2
AC Q9MWJ2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE MHC class IIB antigen (Fragment).
OS Certhidea olivacea.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Passeriformes; Passeroidea;
OC Fringillidae; Emberizinae; Certhidea.
OX NCBI_TaxID=48880;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D1104RF;
RA Sato A., Figueroa F., Mayer W.E., Grant P.R., Grant R., Klein J.;
RT "MHC class II genes of Darwin's finches: divergence by point mutations
  and reciprocal recombination.";
RL (In) Kasahara M. (eds.);
RL Major histocompatibility complex:
  evolution, structure, and function, pp.518-541, Springer-Verlag,
  Tokyo (2000).
DR EMBL; AF164163; AAF36225.1; -.
DR InterPro; IPR000353; MHC II beta.
DR Pfam; PFO0969; MHC II beta; 1.
DR ProDom; PD000328; MHC_II_beta; 1.
FT NON_TER 72 1
FT NON_TER 72 72
SQ SEQUENCE 72 AA; 8760 MW; F62FF27A7B4A672B CRC64;

Query Match 100.0%; Score 21; DB 7; Length 72;
Best Local Similarity 100.0%; Pred. No. 5.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4
Db 63 AAPF 66

RESULT 12
Q26512 PRELIMINARY; PRT; 73 AA.
ID Q26512
AC Q26512;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

```

```

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Triosephosphate isomerase (Fragment).
OS Schistosoma japonicum (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6182;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHINESE AND PHILIPPINE STRAINS;
RX MEDLINE=97101875; PubMed=8946393;
RA Hooker C.W., Brindley P.J.;
RT "Cloning and characterisation of strain-specific transcripts encoding
  triosephosphate isomerase, a candidate vaccine antigen from
  Schistosoma japonicum.";
RL Mol. Biochem. Parasitol. 82:265-269(1996).
DR EMBL; U57557; AAC47394.1; -.
DR HSP; P00940; ITIM.
DR InterPro; IPR000652; Triophos_ismrse.
DR Pfam; PFO0121; TIM; 1.
DR ProDom; PD001005; Triophos_ismrse; 1.
KW Isomerase.
FT NON_TER 1 1
FT NON_TER 73 73
SQ SEQUENCE 73 AA; 7982 MW; 6B10DB35855F5A33 CRC64;

Query Match 100.0%; Score 21; DB 5; Length 73;
Best Local Similarity 100.0%; Pred. No. 5.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4
Db 35 AAPF 38

RESULT 13
Q8UEW5 PRELIMINARY; PRT; 75 AA.
ID Q8UEW5
AC Q8UEW5;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein Atul639.
GN ATU1639.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
Chen Y., Paulsen I.T., Eissen J.A., Karp P.D., Bovee D. Sr.,
Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
Nester E.W.;
RA "The genome of the natural genetic engineer Agrobacterium tumefaciens
  C58.";
RL Science 294:2317-2323(2001).
DR EMBL; AE009121; AAL42640.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 75 AA; 8397 MW; B3B9D20CB8E82DD CRC64;

Query Match 100.0%; Score 21; DB 16; Length 75;
Best Local Similarity 100.0%; Pred. No. 5.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4
Db 1 AAPF 4

```

Db 50 AAPF 53

QY 1 AAPF 4  
 Db 54 AAPF 57

Search completed: February 12, 2003, 10:27:29  
 Job time : 5.82985 secs

## RESULT 14

Q9DFN9 PRELIMINARY; PRT; 76 AA.  
 AC Q9DFN9; 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Formyltetrahydrofolate dehydrogenase (fragment).  
 OS Gillichthys mirabilis (Long-jawed mudsucker).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Gobioidae;  
 OC Gobiidae; Gillichthys.  
 OX NCBI\_TaxID=8222;  
 RN [1]\_TaxID=8222;  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RX MEDLINE=21117151; PubMed=11172064;  
 RA Gracey A.Y., Troll J.V., Somero G.N.;  
 RT "Hypoxia-induced gene expression profiling in the euryoxic fish  
 RT Gillichthys mirabilis."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:1993-1998(2001).  
 DR EMBL; AF266196; AAG13316.1; -.  
 DR HSSP; P05091; LCM3  
 DR InterPro; IPR002086; Aldehyde\_dehydr.  
 DR Pfam; PF00171; algedh; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 76 AA; 8182 MW; 787505071A1ACFA3 CRC64;

Query Match 100.0%; Score 21; DB 13; Length 76;  
 Best Local Similarity 100.0%; Pred. No. 5.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
 Db 43 AAPF 46

## RESULT 15

Q98L67 PRELIMINARY; PRT; 77 AA.  
 AC Q98L67;  
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical protein msr1158.  
 GN MSR1158.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 OX NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NAFF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Matanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti."  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL; AP002996; BAB48596.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 77 AA; 8189 MW; 8A2BF074E4C16806 CRC64;

Query Match 100.0%; Score 21; DB 16; Length 77;  
 Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:18:40 ; Search time 0.859702 Seconds  
(without alignments)  
136.898 Million cell updates/sec

Title: US-10-036-371-7

Perfect score: 21

Sequence: 1 AAPP 4

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA: +  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	4	1	US-07-973-335-1
2	21	100.0	4	1	US-08-153-331-19
3	21	100.0	4	1	US-08-004-643C-2
4	21	100.0	4	1	US-08-462-456-5
5	21	100.0	4	1	US-08-282-860-1
6	21	100.0	4	1	US-08-407-000-6
7	21	100.0	4	1	US-07-890-422B-28
8	21	100.0	4	1	US-08-276-936A-2
9	21	100.0	4	1	US-08-345-820B-2
10	21	100.0	4	1	US-08-544-143A-3
11	21	100.0	4	1	US-08-397-602A-9
12	21	100.0	4	1	US-08-437-029-5
13	21	100.0	4	1	US-08-424-022-19
14	21	100.0	4	1	US-08-439-534-10
15	21	100.0	4	1	US-08-434-959-3
16	21	100.0	4	1	US-08-434-959-4
17	21	100.0	4	1	US-08-434-959-5
18	21	100.0	4	1	US-08-434-959-6
19	21	100.0	4	1	US-08-460-343B-56
20	21	100.0	4	1	US-08-067-180B-2
21	21	100.0	4	1	US-08-398-028B-56
22	21	100.0	4	1	US-08-470-837-3
23	21	100.0	4	2	US-08-722-268-2
24	21	100.0	4	2	US-08-504-265B-56
25	21	100.0	4	2	US-08-439-005-10
26	21	100.0	4	2	US-08-698-575B-2
27	21	100.0	4	2	US-08-424-017B-19

28	21	100.0	4	2	US-08-751-070B-4	Sequence 4, Appli
29	21	100.0	4	3	US-09-069-823-2	Sequence 2, Appli
30	21	100.0	4	3	US-09-076-460-1	Sequence 1, Appli
31	21	100.0	4	3	US-08-950-618-3	Sequence 3, Appli
32	21	100.0	4	3	US-08-950-618-18	Sequence 18, Appli
33	21	100.0	4	3	US-08-950-618-31	Sequence 31, Appli
34	21	100.0	4	3	US-09-120-365-101	Sequence 101, Appl
35	21	100.0	4	3	US-08-319-501-10	Sequence 10, Appl
36	21	100.0	4	4	US-09-168-010-3	Sequence 3, Appli
37	21	100.0	4	4	US-09-168-010-18	Sequence 18, Appl
38	21	100.0	4	4	US-08-905-359A-25	Sequence 25, Appl
39	21	100.0	4	4	US-09-515-039-101	Sequence 101, Appl
40	21	100.0	4	4	US-09-326-039-15	Sequence 15, Appl
41	21	100.0	4	4	US-09-343-650-3	Sequence 3, Appli
42	21	100.0	4	4	US-09-343-650-18	Sequence 18, Appl
43	21	100.0	4	4		
44	21	100.0	4	4		
45	21	100.0	4	4		

#### ALIGNMENTS

RESULT 1  
US-07-973-335-1  
; Sequence 1, Application US/07973335  
; Patent No. 5338547  
; GENERAL INFORMATION:  
; APPLICANT: Kennedy and Szuha  
; TITLE OF INVENTION: No. 5338547el Bowman-Birk Inhibitor  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Woodcock Washburn Kurtz  
; ADDRESSEE: Mackiewicz & No. 5338547ris  
; STREET: One Liberty Place - 46th Floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103

COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07973,335

FILING DATE: 19921102

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 824,719

FILING DATE: January 17, 1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 579,155

FILING DATE: September 6, 1990

ATTORNEY/AGENT INFORMATION:

NAME: Jane Massey Licata

REGISTRATION NUMBER: 32,257

REFERENCE/DOCKET NUMBER: DPSC-0003

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 568-3100

TELEFAX: (215) 568-3439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4

TYPE: AMINO ACID

TOPOLOGY: linear

US-07-973-335-1

Query Match 100.0%; Score 21; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
Db 1 AAPF 4

RESULT 2  
US-08-155-331-19  
; Sequence 19, Application US/08155331  
; Patent No. 5441931  
; GENERAL INFORMATION:  
; APPLICANT: Foster, Donald C  
; APPLICANT: Sprecher, Cindy  
; APPLICANT: No. 5441931is, Kield  
; TITLE OF INVENTION: NOVEL HUMAN AMYLOID PROTEIN PRECURSOR  
; TITLE OF INVENTION: HOMOLOG AND KUNITZ-TYPE INHIBITOR  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ZymoGenetics, Inc.  
; STREET: 4225 Roosevelt Way, N.E.  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/155,331  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 703,590  
; FILING DATE: 20 May 1991  
; APPLICATION NUMBER: 398,092  
; FILING DATE: 24 August 1989  
; APPLICATION NUMBER: DD WP 601 P/319 577W  
; FILING DATE: 07 September 1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Katona, Gabriel P.  
; REGISTRATION NUMBER: 20,829  
; REFERENCE/DOCKET NUMBER: 691-003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)986-3377  
; TELEFAX: (212)986-6126  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acid residues  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-004-643C-2

Query Match 100.0%; Score 21; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2e+05; 0; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
Db 1 AAPF 4

RESULT 4  
US-08-462-456-5  
; Sequence 5, Application US/08462456  
; Patent No. 5510333  
; GENERAL INFORMATION:  
; APPLICANT: Angelastro, Michael R  
; APPLICANT: Bey, Philippe  
; APPLICANT: Doherty, Niall S  
; APPLICANT: Janusz, Michael J  
; APPLICANT: Mehdi, Shujaath  
; APPLICANT: Peet, No. 5510333ton P  
; TITLE OF INVENTION: Inhibitors of Cathepsin G and Elastase  
; TITLE OF INVENTION: for Preventing Connective Tissue Degradation  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marion Merrell Dow Inc.  
; STREET: 2110 East Galbraith Rd.  
; CITY: Cincinnati P. O. Box 156300



```

; STATE: Ohio
; COUNTRY: USA
; ZIP: 45215-6300
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,456
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/342,999
; FILING DATE:
; APPLICATION NUMBER: US/08/222,552
; FILING DATE:
; APPLICATION NUMBER: US/07/987,587
; FILING DATE:
; APPLICATION NUMBER: US/07/704,499
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Nesbitt, Stephen L.
; REGISTRATION NUMBER: 28,981
; REFERENCE/DOCKET NUMBER: M01593
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 948-7965
; TELEFAX: (513) 948-7961
; TELEX: 214320
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-462-456-5

```

```

Query Match 100.0%; Score 21; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 AAPP 4
Db 1 AAPP 4

```

```

RESULT 5
US-08-282-860-1
; Sequence 1, Application US/08282860
; Patent No. 5561108
; GENERAL INFORMATION:
; APPLICANT: Teay, Grace C.
; APPLICANT: Cheung, Neal K. H.
; APPLICANT: Bettencourt, Jeffrey D.
; TITLE OF INVENTION: Preparation of Alpha-1 Antichymotrypsin
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bayer Corporation
; STREET: 800 Dwight Way
; STREET: P. O. Box 1986
; CITY: Berkeley
; STATE: California
; COUNTRY: USA
; ZIP: 94701-1986
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb Storage
; COMPUTER: IBM
; OPERATING SYSTEM: DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/282,860
; FILING DATE: 07/29/94
; CLASSIFICATION: 514

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Giblin, James A.
; REGISTRATION NUMBER: 25772
; REFERENCE/DOCKET NUMBER: MSB-7214
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510)705-7910
; TELEFAX: (510)705-7904
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4
; TYPE: amino acid
; STRANDEDNESS: single strand
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: peptide
; FEATURE:
; NAME/KEY: Modified peptide used for enzyme activity assay
; OTHER INFORMATION: Peptide is modified with n-terminal
; OTHER INFORMATION: succinyl group and c-terminal p-nitroanilide group
; PUBLICATION INFORMATION:
; AUTHORS: Del Mar, E. G.
; AUTHORS: Lagman, C.
; AUTHORS: Broderick, J. W.
; TITLE: A Sensitive New Substrate for Chymotrypsin
; JOURNAL: Analytical Biochemistry
; VOLUME: 99
; PAGES: 316-329
; DATE: 01-NOV-1979
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 4
; US-08-282-860-1

```

```

Query Match 100.0%; Score 21; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 AAPP 4
Db 1 AAPP 4

```

```

RESULT 6
US-08-407-000-6
; Sequence 6, Application US/08407000
; Patent No. 5578324
; GENERAL INFORMATION:
; APPLICANT: Dohi, Masahiko
; APPLICANT: Nishibe, Yoshihisa
; APPLICANT: Makino, Yuji
; APPLICANT: Suzuki, Yoshiki
; TITLE OF INVENTION: PEPTIDE PROTEINACEOUS DRUG NASAL
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/407,000
; FILING DATE: 29-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP94/01257
; FILING DATE: 29-JUL-1994
; APPLICATION NUMBER: JP-A-5-206922

```

; FILING DATE: 30-JUL-1993  
 ; APPLICATION NUMBER: JP-A-5-235841  
 ; FILING DATE: 30-AUG-1993  
 ; APPLICATION NUMBER: JP-A-6-1644  
 ; FILING DATE: 12-JAN-1994  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 4 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; US-08-407-000-6

Query Match 100.0%; Score 21; DB 1; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 2e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
 ||||  
 Db 1 AAPF 4

RESULT 7  
 US-07-890-422B-28  
 ; Sequence 28, Application US/07890422B  
 ; Patent No. 5602102  
 ; GENERAL INFORMATION:  
 ; APPLICANT: THIRLE, DWAIN L.  
 ; APPLICANT: LIPSKY, PETER B.  
 ; APPLICANT: MCGUIRE, MICHAEL J.  
 ; TITLE OF INVENTION: DIPEPTIDYL PEPTIDASE-I  
 ; TITLE OF INVENTION: INHIBITORS AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 38  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: ARNOLD, WHITE & DURKEE  
 ; STREET: P.O. BOX 4433  
 ; CITY: HOUSTON  
 ; STATE: TEXAS  
 ; COUNTRY: USA  
 ; ZIP: 77210

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: FLOPPY DISK  
 ; COMPUTER: IBM PC COMPATIBLE  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: WORDPERFECT 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/890,422B  
 ; FILING DATE: 19920529  
 ; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:  
 ; NAME: MAYFIELD, DENISE L.  
 ; REGISTRATION NUMBER: 33,732  
 ; REFERENCE/DOCKET NUMBER: UTSD:296/MAY  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 512-320-7200  
 ; TELEFAX: 512-474-7577

; TELEX: NOT APPLICABLE  
 ; INFORMATION FOR SEQ ID NO: 28:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 4 amino acid residues  
 ; TYPE: AMINO ACID  
 ; TOPOLOGY: linear

US-07-890-422B-28  
 Query Match 100.0%; Score 21; DB 1; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 2e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
 ||||  
 Db 1 AAPF 4

RESULT 8  
 US-08-276-936A-2  
 ; Sequence 2, Application US/08276936A  
 ; Patent No. 5612194  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Harvey Rubin, Barry Cooperman, No. 5612194man Schecter,  
 ; APPLICANT: Michael Plotkin, Zhi Wang  
 ; TITLE OF INVENTION: Methods of Producing Effective  
 ; TITLE OF INVENTION: Recombinant Serine Protease Inhibitors and Uses of These  
 ; TITLE OF INVENTION: Inhibitors  
 ; NUMBER OF SEQUENCES: 30  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Jane Massey Licata, Esq.  
 ; STREET: 210 Lake Drive East, Suite 201  
 ; CITY: Cherry Hill  
 ; STATE: NJ  
 ; COUNTRY: USA  
 ; ZIP: 08002

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE  
 ; COMPUTER: IBM 486

; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS  
 ; SOFTWARE: WORDPERFECT 5.1

; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/276,936A  
 ; FILING DATE: July 19, 1994

; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 221,078

; FILING DATE: March 31, 1994  
 ; APPLICATION NUMBER: 221,171

; FILING DATE: March 31, 1994  
 ; APPLICATION NUMBER: 005,908

; FILING DATE: January 15, 1993  
 ; APPLICATION NUMBER: 735,335

; FILING DATE: July 24, 1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Jane Massey Licata

; REGISTRATION NUMBER: 32,257  
 ; REFERENCE/DOCKET NUMBER: PENN-0027

; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (609) 779-2400

; TELEFAX: (609) 779-8488  
 ; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 4

; TYPE: Amino Acid  
 ; TOPOLOGY: Linear

US-08-276-936A-2  
 Query Match 100.0%; Score 21; DB 1; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 2e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
 ||||  
 Db 1 AAPF 4

RESULT 9  
 US-08-345-820B-2  
 ; Sequence 2, Application US/08345820B  
 ; Patent No. 5618792  
 ; GENERAL INFORMATION:  
 ; APPLICANT:  
 ; TITLE OF INVENTION: SUBSTITUTED HETEROCYCLIC COMPOUNDS USEFUL AS  
 ; TITLE OF INVENTION: INHIBITORS OF (SERINE PROTEASES) HUMAN NEUTROPHIL ELASTASE  
 ; NUMBER OF SEQUENCES: 4  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/345,820B  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-345-820B-2

Query Match 100.0%; Score 21; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPP 4  
|||  
Db 1 AAPP 4

RESULT 10  
US-08-544-143A-3  
; Sequence 3, Application US/08544143A  
; Patent No. 5646028  
; GENERAL INFORMATION:  
; APPLICANT: Leigh, Scott D.  
; TITLE OF INVENTION: NOVEL ALKALINE PROTEASE AND USES  
; TITLE OF INVENTION: THEREOF  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Majestic, Parsons, Siebert & Hsue  
; STREET: Four Embarcadero Center, Suite 1100  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 94111-4121  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/544,143A  
; FILING DATE: 17-OCT-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: de Runtz, K. Alison  
; REGISTRATION NUMBER: 37,119  
; REFERENCE/DOCKET NUMBER: 0409.054US3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-362-5556  
; TELEFAX: 415-362-5418  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-544-143A-3

Query Match 100.0%; Score 21; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPP 4  
|||  
Db 1 AAPP 4

RESULT 11  
US-08-397-602A-9

; Sequence 9, Application US/08397602A  
; Patent No. 5646044  
; GENERAL INFORMATION:  
; APPLICANT: Wilson, Charles R  
; APPLICANT: Tang, Maria R  
; APPLICANT: Berger, Harald  
; APPLICANT: Christianson, Teresa M  
; APPLICANT: Hansen, Dieter  
; TITLE OF INVENTION: Expression Systems for the Production  
; TITLE OF INVENTION: of Target Proteins in Bacillus  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Henkel Corporation Law Department  
; STREET: 140 Germantown Pike, Suite 150  
; CITY: Plymouth Meeting  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19462  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/397,602A  
; FILING DATE: 02-MAR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jaeschke, Wayne C  
; REGISTRATION NUMBER: 21,062  
; REFERENCE/DOCKET NUMBER: D8969/M4828  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (610) 832-2200  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 1  
; OTHER INFORMATION: /note= "N-SUCCINYL"  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 4  
; OTHER INFORMATION: /note=  
; OTHER INFORMATION: "P-NITROANILIDE"  
US-08-397-602A-9

Query Match 100.0%; Score 21; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPP 4  
|||  
Db 1 AAPP 4

RESULT 12  
US-08-437-029-5  
; Sequence 5, Application US/08437029  
; Patent No. 5666107  
; GENERAL INFORMATION:  
; APPLICANT: Miller, Edward J.  
; TITLE OF INVENTION: Compositions and Methods for  
; TITLE OF INVENTION: Inhibiting Elastase  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew  
; STREET: One Market Plaza, Steuart Tower, Suite 2000  
; CITY: San Francisco

STATE: California  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM: disk  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/437,029  
FILING DATE: 08 May 95  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Dow, Karen B.  
REGISTRATION NUMBER: 29,684  
REFERENCE/DOCKET NUMBER: 017066-000620  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-437-029-5

Query Match 100.0%; Score 21; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
DB 1 AAPF 4

RESULT 13  
US-08-424-022-19  
Sequence 19, Application US/08424022  
Patent No. 5677146  
GENERAL INFORMATION:  
APPLICANT: Foster, Donald C  
APPLICANT: Sprecher, Cindy  
APPLICANT: No. 5677146ris, Kjeld  
TITLE OF INVENTION: NOVEL HUMAN AMYLOID PROTEIN PRECURSOR  
TITLE OF INVENTION: HOMOLOG AND KUNITZ-TYPE INHIBITOR  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ZymoGenetics, Inc.  
STREET: 4225 Roosevelt Way, N.E.  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/424,022  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,692  
FILING DATE: 02-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, Gary E  
REGISTRATION NUMBER: 31-684  
REFERENCE/DOCKET NUMBER: 92-21C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-547-8080 ext 322

TELEFAX: 206-548-2329  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1..4  
OTHER INFORMATION: /label= ALA-1  
OTHER INFORMATION: /note= "Amino terminal alanine residue is capped  
OTHER INFORMATION: with a succinyl group"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1..4  
OTHER INFORMATION: /label= Phe-4  
OTHER INFORMATION: /note= "Carboxyl-terminal phenylalanine residue is  
OTHER INFORMATION: capped with p-nitroanil..."  
US-08-424-022-19

Query Match 100.0%; Score 21; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
DB 1 AAPF 4

RESULT 14  
US-08-439-534-10  
Sequence 10, Application US/08439534  
Patent No. 5719041  
GENERAL INFORMATION:  
APPLICANT: Lazarus, Robert A.  
APPLICANT: Dennis, Mark S.  
APPLICANT: Ulmer, Jana S.  
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTAINING  
TITLE OF INVENTION: ECOTIN AND HOMOLOGS THEREOF  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/439,534  
FILING DATE: 11-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/319501  
FILING DATE: 04-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/121004  
FILING DATE: 14-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Kubinec, Jeffrey S.  
REGISTRATION NUMBER: 36,575  
REFERENCE/DOCKET NUMBER: P0859C1D2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-8228  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-439-534-10

Query Match 100.0%; Score 21; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred No. 2e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAPF 4  
Db 1 AAPF 4

RESULT 15  
US-08-434-959-3  
Sequence 3, Application US/08434959  
Patent No. 5736520  
GENERAL INFORMATION:  
APPLICANT: Bey, Philippe  
APPLICANT: Angelastro, Michael R  
APPLICANT: Mehdi, Shujaath  
TITLE OF INVENTION: No. 5736520el Peptidase Inhibitors  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marion Merrell Dow Inc.  
STREET: 2110 East Galbraith Rd.  
CITY: Cincinnati P. O. Box 156300  
STATE: Ohio  
COUNTRY: USA  
ZIP: 45215-6300  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/434,959  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/750,439  
FILING DATE: 20-AUG-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/454,803  
FILING DATE: 21-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/439,201  
FILING DATE: 20-NOV-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/416,817  
FILING DATE: 04-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/254,762  
FILING DATE: 07-OCT-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Nesbitt, Stephen L.  
REGISTRATION NUMBER: 28,981  
REFERENCE/DOCKET NUMBER: M01368F US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (513) 948-7965  
TELEFAX: (513) 948-7961  
TELEX: 214320  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-434-959-3

Query Match 100.0%; Score 21; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAPF 4  
Db 1 AAPF 4

Search completed: February 12, 2003, 10:30:11  
Job time : 1.8597 secs

**THIS PAGE BLANK (USPTO)**



; PRIOR APPLICATION NUMBER: US 09/326,039  
; PRIOR FILING DATE: 1999-06-04  
; PRIOR APPLICATION NUMBER: US 60/088,136  
; PRIOR FILING DATE: 1998-06-04  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 15  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-757-908A-15

Query Match 100.0%; Score 21; DB 10; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.2e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
Db 1 AAPF 4

## RESULT 3

US-09-117-380B-2  
; Sequence 2, Application US/09117380B  
; Patent No. US20020119917A1  
; GENERAL INFORMATION:  
; APPLICANT: FRIDKIN, Matityahu  
; APPLICANT: YAVIN, Eran J.  
; TITLE OF INVENTION: ANTI-INFLAMMATORY PEPTIDES DERIVED FROM C-REACTIVE  
; FILE REFERENCE: FRIDKIN=1  
; CURRENT APPLICATION NUMBER: US/09/117,380B  
; PRIOR FILING DATE: 1999-01-27  
; PRIOR APPLICATION NUMBER: PCT/IL97/00032  
; PRIOR FILING DATE: 1997-01-27  
; PRIOR APPLICATION NUMBER: IL 116976  
; PRIOR FILING DATE: 1996-01-31  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; OTHER INFORMATION: The N-terminal Ala residue is modified with a  
; OTHER INFORMATION: succinyl group; the C-terminal Phe residue is  
; OTHER INFORMATION: modified with a nitroanilide group.  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-117-380B-2

Query Match 100.0%; Score 21; DB 10; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.2e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
Db 1 AAPF 4

## RESULT 4

US-09-935-744-1  
; Sequence 1, Application US/09935744  
; Patent No. US20020137118A1  
; GENERAL INFORMATION:  
; APPLICANT: Inouye, Masayori  
; APPLICANT: Shinde, Ujwal  
; APPLICANT: Fu, Xuan  
; TITLE OF INVENTION: Biologically Active Protein Folding Intermediates  
; FILE REFERENCE: 266/223  
; CURRENT APPLICATION NUMBER: US/09/935,744  
; CURRENT FILING DATE: 2001-08-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patentin version 3.1

; SEQ ID NO 1  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: The sequence is part of a synthetic peptide that is used as a sub  
; OTHER INFORMATION: strate for determining the activation time of a stable crosslinke  
; OTHER INFORMATION: d intermediate conformer.  
; NAME/KEY: MISC FEATURE  
; LOCATION: (1)-(4)  
; OTHER INFORMATION: Sequence is preceded by N-succinyl and followed by p-nitroanilide  
US-09-935-744-1

Query Match 100.0%; Score 21; DB 10; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.2e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
Db 1 AAPF 4

## RESULT 5

US-10-090-624-33  
; Sequence 33, Application US/10090624  
; Patent No. US2002013235A1  
; GENERAL INFORMATION:  
; APPLICANT: TAKAKURA, Hikaru  
; APPLICANT: MORISHITA, Mio  
; APPLICANT: SHIMOJO, Tomoko  
; APPLICANT: ASADA, Kiyozo  
; APPLICANT: KATO, Kunoshin  
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE  
; FILE REFERENCE: TAKAKURA=6  
; CURRENT APPLICATION NUMBER: US/10/090,624  
; CURRENT FILING DATE: 2002-03-06  
; PRIOR APPLICATION NUMBER: 09/445,472  
; PRIOR FILING DATE: 1999-12-06  
; PRIOR APPLICATION NUMBER: 151969/1997  
; PRIOR FILING DATE: 1997-06-10  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 33  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Residue 1 is modified by a succinyl group.  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Residue 4 is modified by a p-nitroaniline group.  
US-10-090-624-33

Query Match 100.0%; Score 21; DB 12; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.2e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
Db 1 AAPF 4

## RESULT 6

US-10-036-371-7  
; Sequence 7, Application US/10036371  
; Patent No. US20020141987A1  
; GENERAL INFORMATION:  
; APPLICANT: BJARNARSON, JON B.  
; TITLE OF INVENTION: FISH SERINE PROTEINASES AND THEIR PHARMACEUTICAL AND  
; SOFTWARE: COSMETIC USE



```
; FILE REFERENCE: 81691/284960
; CURRENT APPLICATION NUMBER: US/10/036,371
; CURRENT FILING DATE: 2002-01-07
; PRIOR FILING DATE: 09/411,688
; PRIOR FILING DATE: 1999-10-12
; PRIOR APPLICATION NUMBER: 5086/99
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic peptide
US-10-036-371-7
```

```
Query Match 100.0%; Score 21; DB 12; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAPF 4
Db 1 AAPF 4
```

```
RESULT 7
US-10-040-655-10
; Sequence 10, Application US/10040655
; Patent No. US20020146805A1
; GENERAL INFORMATION:
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Darrow, Andrew
; APPLICANT: Qi, Jian-shen
; TITLE OF INVENTION: DNA encoding the novel human serine
; FILE OF INVENTION: Protease T
; FILE REFERENCE: ORT-1032
; CURRENT APPLICATION NUMBER: US/10/040,655
; CURRENT FILING DATE: 2002-01-07
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: peptide substrate
US-10-040-655-10
```

```
Query Match 100.0%; Score 21; DB 12; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAPF 4
Db 1 AAPF 4
```

```
RESULT 8
US-10-033-526-1
; Sequence 1, Application US/10033526
; Patent No. US20020147999A1
; GENERAL INFORMATION:
; APPLICANT: Robert W. Mahley
; APPLICANT: Yadong Huang
; TITLE OF INVENTION: Methods of Treating Disorders Related to
; FILE OF INVENTION: APOE
; FILE REFERENCE: UCAL217
; CURRENT APPLICATION NUMBER: US/10/033,526
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 60/245,737
; PRIOR FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 4
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic peptide
US-10-033-526-1
```

```
Query Match 100.0%; Score 21; DB 12; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAPF 4
Db 1 AAPF 4
```

```
RESULT 9
US-09-994-927-1
; Sequence 1, Application US/09994927
; Patent No. US20020127605A1
; GENERAL INFORMATION:
; APPLICANT: Guilford Pharmaceuticals Inc.
; APPLICANT: Hamilton, Gregory
; APPLICANT: Belyakov, Sergei
; APPLICANT: Vaal, Mark
; APPLICANT: Wei, Ling
; APPLICANT: Wu, Yong-Qian
; APPLICANT: Steiner, Joseph
; TITLE OF INVENTION: Bisubstituted Carbocyclic Cyclophilin Binding Compounds and Their
; FILE REFERENCE: 03166.0029, NEUS02
; CURRENT APPLICATION NUMBER: US/09/994,927
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 60/253,074
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 60/291,966
; PRIOR FILING DATE: 2001-05-21
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Model substrate for measuring rotamase inhibition activity
; NAME/KEY: MISC FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa at position 1 is N-succinyl
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: Xaa at position 6 is p-nitroanilide
US-09-994-927-1
```

```
Query Match 100.0%; Score 21; DB 10; Length 6;
Best Local Similarity 100.0%; Pred. No. 1.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAPF 4
Db 2 AAPF 5
```

```
RESULT 10
US-10-033-526-4
; Sequence 4, Application US/10033526
; Patent No. US20020147999A1
; GENERAL INFORMATION:
; APPLICANT: Robert W. Mahley
; APPLICANT: Yadong Huang
; TITLE OF INVENTION: Methods of Treating Disorders Related to
; FILE OF INVENTION: APOE
; FILE REFERENCE: UCAL217
```

1	PRIOR APPLICATION NUMBER:	PCT/US01/006651
2	PRIOR FILING DATE:	2001-01-30
3	PRIOR APPLICATION NUMBER:	PCT/US01/006707
4	PRIOR FILING DATE:	2001-01-30
5	PRIOR APPLICATION NUMBER:	US 60/234,687
6	PRIOR FILING DATE:	2000-09-21
7	PRIOR APPLICATION NUMBER:	US 09/608,408
8	PRIOR FILING DATE:	2000-06-30
9	PRIOR APPLICATION NUMBER:	US 09/774,203

;  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 38169  
; LENGTH: 25  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL050342.36  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.91  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.95  
US-09-864-761-38169

Query Match 100.0%; Score 21; DB 10; Length 25;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
|||  
Db 2 AAPF 5

RESULT 14  
US-09-974-879-335  
; Sequence 335, Application US/09974879  
; Publication No. US20030028003A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 125 Human Secreted Proteins  
; FILE REFERENCE: P2020P2  
; CURRENT APPLICATION NUMBER: US 09/974,879  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: US 60/239,893  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: US 09/818,683  
; PRIOR FILING DATE: 2001-03-28  
; PRIOR APPLICATION NUMBER: US 09/305,736  
; PRIOR FILING DATE: 1999-05-05  
; PRIOR APPLICATION NUMBER: PCT/US98/23435  
; PRIOR FILING DATE: 1998-11-04  
; PRIOR APPLICATION NUMBER: US 60/064,911  
; PRIOR FILING DATE: 1997-11-07  
; PRIOR APPLICATION NUMBER: US 60/064,912  
; PRIOR FILING DATE: 1997-11-07  
; PRIOR APPLICATION NUMBER: US 60/064,983  
; PRIOR FILING DATE: 1997-11-07  
; PRIOR APPLICATION NUMBER: US 60/064,900  
; PRIOR FILING DATE: 1997-11-07  
; PRIOR APPLICATION NUMBER: US 60/064,988  
; PRIOR FILING DATE: 1997-11-07  
; PRIOR APPLICATION NUMBER: US 60/064,987  
; PRIOR FILING DATE: 1997-11-07  
; PRIOR APPLICATION NUMBER: US 60/064,908  
; PRIOR FILING DATE: 1997-11-07  
; PRIOR APPLICATION NUMBER: US 60/064,984  
; PRIOR FILING DATE: 1997-11-07  
; PRIOR APPLICATION NUMBER: US 60/064,985  
; PRIOR FILING DATE: 1997-11-07  
; PRIOR APPLICATION NUMBER: US 60/066,094  
; PRIOR FILING DATE: 1997-11-17  
; PRIOR APPLICATION NUMBER: US 60/066,100  
; PRIOR FILING DATE: 1997-11-17  
; PRIOR APPLICATION NUMBER: US 60/066,089  
; PRIOR FILING DATE: 1997-11-17  
; PRIOR APPLICATION NUMBER: US 60/066,095  
; PRIOR FILING DATE: 1997-11-17  
; PRIOR APPLICATION NUMBER: US 60/066,090  
; PRIOR FILING DATE: 1997-11-17  
; NUMBER OF SEQ ID NOS: 611  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 335  
; LENGTH: 31  
; TYPE: PRT

;  
; ORGANISM: Homo sapiens  
US-09-974-879-335

Query Match 100.0%; Score 21; DB 9; Length 31;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
|||  
Db 20 AAPF 23

RESULT 15  
US-09-864-761-48747  
; Sequence 48747, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aseonix-X-1  
; CURRENT APPLICATION NUMBER: US 09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 48747  
; LENGTH: 45  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC006389.2  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.8  
; OTHER INFORMATION: SWISSPROT HIT: P15772, EVALU8 8.20e+00

US-09-864-761-48747

Query Match 100.0%; Score 21; DB 10; Length 45;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAPF 4  
|  
|  
|  
|  
Db 39 AAPF 42

Search completed: February 12, 2003, 10:31:07  
Job time : 1.60896 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:03:55 ; Search time 13.4328 Seconds  
(without alignments)  
198.395 Million cell updates/sec

Title: US-10-036-371-1

Perfect score: 95

Sequence: 1 IVGGYXCKXHSQAHQVSLNS 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002.\*  
1: /SID32/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID32/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID32/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID32/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID32/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID32/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID32/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID32/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID32/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID32/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID32/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID32/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID32/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID32/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID32/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID32/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID32/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID32/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID32/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID32/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID32/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID32/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID32/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	90	94.7	23	21	AA199397
2	90	94.7	23	22	AAE07943
3	90	94.7	23	22	AAE07944
4	90	94.7	25	20	AA193347
5	90	94.7	25	20	AA193348
6	90	94.7	25	21	AA192966
7	90	94.7	25	21	AA192967
8	90	94.7	25	23	AAO21358
9	90	94.7	25	23	AAO21359
10	90	94.7	37	21	AA193946

11	90	94.7	37	23	AAO21360	Atlantic cod multi
12	90	94.7	225	22	AA191579	Amino acid sequenc
13	79	83.2	240	23	AB046444	Engraulis japonicu
14	78	82.1	25	20	AA193346	Atlantic cod hydro
15	78	82.1	25	21	AA192965	Salmon enzyme I mu
16	78	82.1	25	21	AA193336	N-terminal of an a
17	78	82.1	25	22	AAE07942	Salmon enzyme I mu
18	78	82.1	25	23	ABO21357	Engraulis japonicu
19	77	81.1	241	23	ABO46445	Human amyloid beta
20	71	74.7	246	19	AAW64260	Human terminal seq
21	68.5	72.1	19	22	AA191573	Amino terminal seq
22	66	69.5	20	21	AA197897	Canine anionic try
23	66	69.5	246	21	AA197897	Canine anionic try
24	66	69.5	247	18	AAW08475	Porcine trypsinoge
25	65	68.4	240	19	AAW57740	Trypsinogen-like p
26	65	68.4	247	9	AA191243	Human spleen tryps
27	65	68.4	247	16	AA192703	Human pancreatic t
28	64	67.4	224	20	AA191160	Human trypsin seri
29	64	67.4	225	22	AA198503	Human TRYI trypsin
30	64	67.4	230	20	AAW93488	Human trypsinogen
31	64	67.4	241	21	AA191316	Human trypsinogen
32	64	67.4	247	21	AA191321	Human trypsinogen
33	64	67.4	247	23	AAU67693	Human pancreatic t
34	64	67.4	254	23	AAU74758	Human pancreatic t
35	63	66.3	20	21	AA190386	N-terminus of bovi
36	63	66.3	20	22	AA191574	Amino terminal seq
37	63	66.3	223	20	AAW81767	Bovine TRYP peptid
38	63	66.3	223	21	AA196973	TRYP protein. Bos
39	63	66.3	224	15	AA195367	Bovine trypsin. B
40	63	66.3	229	21	AA197494	Bovine trypsinogen
41	63	66.3	230	15	AA195368	Bovine trypsinogen
42	63	66.3	231	21	AA191926	Recombinant trypsi
43	63	66.3	231	22	AA198053	Bovine met-phe-try
44	63	66.3	233	21	AA191925	Trypsinogen analog
45	62	65.3	223	23	AA193322	Partial trypsin se

## ALIGNMENTS

RESULT 1  
AA193937  
ID AA193937 standard; peptide; 23 AA.

XX AA193937;

AC XX

DT 03-OCT-2000 (first entry)

XX XX

DE N-terminal of an atlantic cod hydrolase enzyme.

XX XX

KW Transplantation rejection; hydrolase; graft versus host disease;

KW cell surface adhesion molecule; immune reaction; inflammation; shock;

KW tumour metastasis; autoimmune disease;

KW Krill derived multifunctional enzyme.

XX Gadus sp.

XX WO200038708-A1.

XX 06-JUL-2000.

XX 23-DEC-1999; 99WO-US30818.

XX 24-DEC-1998; 98US-0114147.

XX (PHAI-) PHAIRSON MEDICAL INC.

XX Franklin RL, St Pierre Y;

XX WPI; 2000-452301/39.

XX Preventing or ameliorating transplantation rejection reactions using

PT hydrolase enzymes -

PS Disclosure; Page 26; 66pp; English.

XX The specification describes a method for preventing or ameliorating

CC transplantation rejection reactions for transplantation of immune cells

CC or other tissues. The method comprises treating a source of immune cells

CC with a hydrolase or hydrolase mixture and administering the treated

CC immune cells to a recipient animal. The hydrolase especially has a

CC preference for removing, destroying, inactivating or disabling at least

CC one of CD4, CD8, CD28, ICAM-1 (CD54), CD152, an integrin, CD134,

CC CD40 and CD80 in contrast to removing, destroying, inactivating or

CC disabling TcR. The methods are useful for preventing graft versus host

CC disease by using hydrolase enzymes to remove the cell surface adhesion

CC molecules which are involved in triggering the immune reactions involved

CC in the diseases. The methods are used for treating or preventing

CC cell-cell or cell-virus adhesion syndrome comprising inflammation, shock,

CC tumour metastases, autoimmune disease, transplantation rejection

CC reactions or microbial infections. The present sequence represents the

CC N-terminal of a hydrolase, which may be used in the method of the

XX invention.

SQ Sequence 23 AA;

Query Match 94.7%; Score 90; DB 21; Length 23;

Best Local Similarity 85.0%; Pred. No. 1.2e-08;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 IVGGYXCXXHSQAHQVSLNS 20

Db 1 IVGGYECTKHSQAHQVSLNS 20

RESULT 2

ID AAE07943

XX AAE07943 standard; peptide; 23 AA.

AC AAE07943;

XX 01-NOV-2001 (first entry)

XX N-terminal of atlantic cod I serine multifunctional hydrolase.

XX Multifunctional hydrolase; rejection reaction; extra-corporeal;

KW therapy; graft versus host disease; transplantation rejection;

KW autoimmune disease; microbial infection; immune disorder; cytostatic;

KW cystic fibrosis; chronic obstructive pulmonary disease; COPD;

KW atherosclerosis; cancer; asthma; septic shock; toxic shock syndrome;

KW conjunctivitis; reperfusion injury; pain; immunosuppressive;

KW antibacterial; vasotropic; atlantic cod.

OS Gadus morhua I.

XX US6232088-B1.

XX 15-MAY-2001.

XX 24-DEC-1998; 98US-0220731.

XX 08-FEB-1995; 95US-0385540.

XX 07-JUN-1995; 95US-0486820.

XX 08-FEB-1996; 96US-0600273.

XX (PHAI-) PHAIRSON MEDICAL INC.

XX Franklin RL, St Pierre Y;

XX WPI; 2001-450051/48.

XX Ameliorating, preventing or treating immune rejection reactions, such

PT as graft versus host disease, autoimmune disease, asthma, cancer, by

XX extra-corporeally treating donor tissue with hydrolase such as a

PT protease

XX

PS Disclosure; Column 15-16; 27pp; English.

XX The present invention relates to a method for ameliorating

CC transplantation rejection reaction. The method comprises treating

CC extra-corporeally donor tissue or donor source of immune cells with

CC a rejection ameliorating effective amount of a hydrolase enzyme. The

CC method is useful for ameliorating, treating or preventing immune

CC rejection reactions such as graft versus host diseases, organ or tissue

CC transplantation rejection, autoimmune disease and associated conditions,

CC microbial infection, immune disorder, cystic fibrosis, chronic

CC obstructive pulmonary disease (COPD), atherosclerosis, cancer, asthma,

CC septic shock, toxic shock syndrome, conjunctivitis, reperfusion injury

CC and pain in humans. The present sequence is the N-terminal peptide

CC of atlantic cod I serine multifunctional hydrolase, used in the

XX exemplification of the invention.

SQ Sequence 23 AA;

Query Match 94.7%; Score 90; DB 22; Length 23;

Best Local Similarity 85.0%; Pred. No. 1.2e-08;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 IVGGYXCXXHSQAHQVSLNS 20

Db 1 IVGGYECTKHSQAHQVSLNS 20

RESULT 3

ID AAE07944

XX AAE07944 standard; peptide; 23 AA.

AC AAE07944;

XX 01-NOV-2001 (first entry)

XX N-terminal of atlantic cod II serine multifunctional hydrolase.

XX Multifunctional hydrolase; rejection reaction; extra-corporeal;

KW therapy; graft versus host disease; transplantation rejection;

KW autoimmune disease; microbial infection; immune disorder; cytostatic;

KW cystic fibrosis; chronic obstructive pulmonary disease; COPD;

KW atherosclerosis; cancer; asthma; septic shock; toxic shock syndrome;

KW conjunctivitis; reperfusion injury; pain; immunosuppressive;

KW antibacterial; vasotropic; atlantic cod.

OS Gadus morhua II.

XX US6232088-B1.

XX 15-MAY-2001.

XX 24-DEC-1998; 98US-0220731.

XX 08-FEB-1995; 95US-0385540.

XX 07-JUN-1995; 95US-0486820.

XX 08-FEB-1996; 96US-0600273.

XX (PHAI-) PHAIRSON MEDICAL INC.

XX Franklin RL, St Pierre Y;

XX WPI; 2001-450051/48.

XX Ameliorating, preventing or treating immune rejection reactions, such

PT as graft versus host disease, autoimmune disease, asthma, cancer, by

XX extra-corporeally treating donor tissue with hydrolase such as a

PT protease

XX

PS Disclosure; Column 15-16; 27pp; English.

XX The present invention relates to a method for ameliorating

CC transplantation rejection reaction. The method comprises treating

CC extra-corporeally donor tissue or donor source of immune cells with

CC a rejection ameliorating effective amount of a hydrolase enzyme. The  
 CC method is useful for ameliorating, treating or preventing immune  
 CC rejection reactions such as graft versus host diseases, organ or tissue  
 CC transplantation rejection, autoimmune disease and associated conditions,  
 CC microbial infection, immune disorder, cystic fibrosis, chronic  
 CC obstructive pulmonary disease (COPD), atherosclerosis, cancer, asthma,  
 CC septic shock, toxic shock syndrome, conjunctivitis, reperfusion injury  
 CC and pain in humans. The present sequence is the N-terminal peptide  
 CC of atlantic cod II serine multifunctional hydrolase, used in the  
 CC exemplification of the invention.  
 CC Note: The sequence is stated as being the same as that shown as SEQ  
 CC ID NO 19 (see AAE07951) in column 31-32 of the specification. However  
 CC the sequences differ at several positions.  
 XX  
 XX

SQ Sequence 23 AA;

Query Match 94.7%; Score 90; DB 22; Length 23;  
 Best Local Similarity 85.0%; Pred. No. 1.2e-08;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 IVGGYXCXHXSHQHQVSLNS 20  
 Db 1 IVGGYECTRHSHQHQVSLNS 20  
 ||||| | ||||| |||||

#### RESULT 4

AA333347  
 ID AAY33347 standard; peptide; 25 AA.

XX AC AAY33347;

XX DT 29-NOV-1999 (first entry)

XX DE Atlantic cod hydrolase N-terminal fragment II. \*

XX KW Multifunctional enzyme; treatment; acne; eczema; chymotrypsin; trypsin;  
 KW collagenase; elastase; exopeptidase; cell surface receptor; anti-acne;  
 KW antiseborrheic; dermatological; anti-eczema; adhesion molecule; ICAM-1;  
 KW CD54; ICAM-2; VCAM-1; CD4; CD8; CD28; CD29D; CD31; CD44; CD49; CD62L;  
 KW CD102; GM1 ceramide; krill; atlantic cod.

XX OS Gadus sp.

XX US95958406-A.

XX PD 28-SEP-1999

XX PF 08-FEB-1996; 96US-0600273.

XX PR 22-NOV-1994; 94US-0388501.

XX PR 08-FEB-1995; 95US-0385540.

XX PR 07-JUN-1995; 95US-0486820.

XX PA (PHAI-) PHAIRSON MEDICAL INC.

XX PI Lindblom R, Kay J, Franklin RL, De Faire JR;

XX WPI; 1999-561004/47.

XX PT Treating acne and eczema using a krill-derived multifunctional enzyme -  
 XX Disclosure; Column 21-22; 42pp; English.

XX PS This invention describes a novel method for treating acne and eczema  
 CC using a krill-derived multifunctional enzyme (I) which comprises 2 or  
 CC more of the activities of chymotrypsin, trypsin, collagenase, elastase  
 CC or exopeptidase and is reactive with cell surface receptors such as  
 CC proteins or glycoproteins. The product of the invention have  
 CC antiseborrheic, anti-acne, dermatological and anti-eczema activity.  
 CC (I) removes or inactivates cell surface receptors (proteins and  
 CC glycoproteins) and adhesion molecules such as ICAM-1 (i.e. CD54), ICAM-2,  
 CC VCAM-1, CD4, CD8, CD28, CD29D, CD31, CD44, CD49, CD62L CD102 and the  
 CC asialo GM1 ceramide. This sequence represents the N-terminal fragment of

CC an atlantic cod (Gadus sp.) hydrolase which is used to describe the  
 CC method of the invention.

XX SQ Sequence 25 AA;

Query Match 94.7%; Score 90; DB 20; Length 25;  
 Best Local Similarity 85.0%; Pred. No. 1.3e-08;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 IVGGYXCXHXSHQHQVSLNS 20  
 Db 1 IVGGYECTRHSHQHQVSLNS 20  
 ||||| | ||||| |||||

#### RESULT 5

AA333348  
 ID AAY33348 standard; peptide; 25 AA.

XX AC AAY33348;

XX DT 29-NOV-1999 (first entry)

XX DE Krill-derived multifunctional enzyme N-terminal fragment 3.

XX KW Multifunctional enzyme; treatment; acne; eczema; chymotrypsin; trypsin;  
 KW collagenase; elastase; exopeptidase; cell surface receptor; anti-acne;  
 KW antiseborrheic; dermatological; anti-eczema; adhesion molecule; ICAM-1;  
 KW CD54; ICAM-2; VCAM-1; CD4; CD8; CD28; CD29D; CD31; CD44; CD49; CD62L;  
 KW CD102; GM1 ceramide; krill.

XX OS Euphasia sp.

XX US95958406-A.

XX PD 28-SEP-1999.

XX PR 08-FEB-1996; 96US-0600273.

XX PR 22-NOV-1994; 94US-0388501.

XX PR 08-FEB-1995; 95US-0385540.

XX PR 07-JUN-1995; 95US-0486820.

XX PA (PHAI-) PHAIRSON MEDICAL INC.

XX PI Lindblom R, Kay J, Franklin RL, De Faire JR;

XX WPI; 1999-561004/47.

XX PT Treating acne and eczema using a krill-derived multifunctional enzyme -  
 XX Claim 1; Column 57-58; 42pp; English.

XX PS This invention describes a novel method for treating acne and eczema  
 CC using a krill-derived multifunctional enzyme (I) which comprises 2 or  
 CC more of the activities of chymotrypsin, trypsin, collagenase, elastase  
 CC or exopeptidase and is reactive with cell surface receptors such as  
 CC proteins or glycoproteins. The product of the invention have  
 CC antiseborrheic, anti-acne, dermatological and anti-eczema activity.  
 CC (I) removes or inactivates cell surface receptors (proteins and  
 CC glycoproteins) and adhesion molecules such as ICAM-1 (i.e. CD54), ICAM-2,  
 CC VCAM-1, CD4, CD8, CD28, CD29D, CD31, CD44, CD49, CD62L CD102 and the  
 CC asialo GM1 ceramide. This invention describes the N-terminal fragment of  
 CC a krill multifunctional enzyme which is used in the method of the  
 CC invention.

SQ Sequence 25 AA;

Query Match 94.7%; Score 90; DB 20; Length 25;  
 Best Local Similarity 85.0%; Pred. No. 1.3e-08;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 IVGGYXCXHXSHQHQVSLNS 20  
 ||||| | ||||| |||||

Wed Feb 12 11:59:30 2003

us-10-036-371-1.rag

Db	1	IVGGYECTRHSHQHVSLNS	20
RESULT 6			
ID	AAB22966	standard; peptide; 25 AA.	
XX	AC	AAB22966;	
XX	DT	10-JAN-2001 (first entry)	
XX	DE	Atlantic cod multifunctional hydrolase I peptide.	
XX	KW	Multifunctional hydrolase; multifunctional protein;	
XX	KW	protease activity; chymotrypsin; trypsin; collagenase;	
XX	KW	elastase; wound healing; corneal ulcer; internal trauma;	
XX	KW	internal surgical wound; nerve; tendon; sheath;	
XX	KW	adhesion formation inhibition; ophthalmological; vulnery.	
XX	OS	Gadus morhua.	
XX	XX	WO200049991-A2.	
XX	XX	31-AUG-2000.	
XX	PF	29-JUN-1999; 99WO-US14751.	
XX	PR	23-FEB-1999; 99US-0256484.	
XX	XX	(PHAI-) PHAIRSON MEDICAL INC.	
XX	XX	Franklin R, Cowling D, Hubbell JA;	
XX	XX	WPI; 2000-587120/55.	
XX	DR	Use of microgel comprising crosslinked polyanionic polymer and	
XX	PT	optionally proteolytic enzyme in treatment of corneal wounds e.g.	
XX	PT	ulcers and abrasions and internal trauma e.g. surgical wounds and for	
XX	PT	treating implants to reduce adhesions -	
XX	PS	Disclosure; Page 25; 85pp; English.	
XX	XX	The invention relates to use of a microgel comprising a crosslinked	
XX	CC	polyanionic polymer in compositions for treatment of an area affected by	
XX	CC	a trauma selected from corneal wounds and internal trauma. The	
XX	CC	compositions further comprise a multifunctional protease having	
XX	CC	activities selected from two of chymotrypsin activity, trypsin activity,	
XX	CC	collagenase activity and elastase activity. In particular, the protease	
XX	CC	may be one of six isoforms of white shrimp (Penaeus vannamei)	
XX	CC	multifunctional hydrolase (AAB22947-B22952), or a multifunctional enzyme	
XX	CC	comprising one of the peptide sequences given in AAB22953-B22967.	
XX	CC	Compositions of the invention may be used to treat corneal ulcers	
XX	CC	(including infected ulcers), or abrasions or a chemical or physical	
XX	CC	insult to the cornea that are likely to give rise to a corneal ulcer.	
XX	CC	They are also used to treat internal traumas selected from an	
XX	CC	internal surgical wound, or a trauma to a membrane that covers either	
XX	CC	an internal organ or tissue or the cavity in which one or more internal	
XX	CC	organs or tissues reside, with the membrane being selected from the	
XX	CC	peritoneum, the pericardium, the epicardium and the pleura and a	
XX	CC	meninges. The internal trauma is to a tendon, tendon sheath, a nerve	
XX	CC	or a nerve sheath, where the trauma is susceptible to giving rise to	
XX	CC	adhesions. The amount of microgel administered is sufficient to prevent	
XX	CC	or reduce formation or reformation of adhesions, and the composition can	
XX	CC	also be used to treat an implant to reduce formation of adhesions.	
XX	CC	Sequences AAB22953-B22967 represent peptide sequences from a variety of	
XX	CC	Multifunctional hydrolases. Multifunctional hydrolases comprising one of	
XX	CC	these sequences are claimed for use in compositions of the invention.	
XX	SQ	Sequence 25 AA;	

  

Query Match	94.7%;	Score 90;	DB 21;	Length 25;
Best Local Similarity	85.0%;	Pred. No. 1.3e-08;		
Matches 17;	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0;



Query Match 94.7%; Score 90; DB 21; Length 25;  
 Best Local Similarity 85.0%; Pred. No. 1.3e-08;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYKXHXSHQAHQVSLNS 20  
 ||||| | ||||| |||||  
 DB 1 IVGGYECTRHSHQAHQVSLNS 20  
 ||||| | ||||| |||||

RESULT 8  
 AAO21358  
 ID AAO21358 standard; Peptide; 25 AA.  
 AC AAO21358;  
 XX  
 XX 05-AUG-2002 (first entry)  
 XX Atlantic cod I multifunctional hydrolase peptide region.  
 KW Tranquilliser; vulnary; cytostatic; gynaecological; antiinflammatory;  
 KW corneal wound; internal trauma; polyanionic polymer; corneal ulcer;  
 KW adhesion; cervical spondylosis; organ membrane trauma; tissue; cavity;  
 KW endometriosis; pelvic inflammatory disease; cumulative trauma disorder; appendicitis;  
 KW peritonitis; pericarditis; pleuritis; inflammatory disease; adhesive peritonitis;  
 KW multifunctional protein; hydrolase; enzyme; atlantic cod.  
 OS Gadus morhua.  
 XX  
 PN WO200215913-A1.  
 XX  
 PD 28-FEB-2002.  
 XX  
 PF 23-AUG-2000; 2000WO-US23072.  
 XX  
 PR 23-AUG-2000; 2000WO-US23072.  
 XX  
 PA (PHAT-) PHAIRSON MEDICAL INC.  
 PA (EIDG-) EIDGENOSSISCHE TECH HOCHSCHULE ZUERICH.  
 XX  
 PI Franklin R, Cowling DSP, Hubbell JA, Van De Wetering P;  
 XX  
 DR WPI; 2002-425632/45.  
 XX  
 PT Treatment of corneal wounds, internal trauma or inflammatory diseases  
 PT comprises administration of a polyanionic polymer -  
 XX  
 PS Disclosure; Page 24; 59pp; English.  
 XX  
 CC The invention relates to a method for treating corneal wounds or internal  
 CC trauma. The method provides administration of a composition comprising a  
 CC polyanionic polymer. The method is useful for treating corneal wounds,  
 CC including corneal ulcers, corneal abrasions or injuries which may give  
 CC rise to corneal ulcers. The method is also useful for treating internal  
 CC trauma, including surgical wounds, trauma to a membrane that covers an  
 CC organ, tissue or cavity or trauma likely to give rise to adhesions. A  
 CC further method is useful for treating cervical spondylosis, cumulative  
 CC trauma disorder, endometriosis, pelvic inflammatory disease, adhesive  
 CC peritonitis, appendicitis, pericarditis, pleuritis, pericarditis and pleuritis or  
 CC other inflammatory diseases. This sequence represents the atlantic cod I  
 CC multifunctional hydrolase peptide region relating to the polyanionic  
 CC polymers of the invention.  
 XX  
 SQ Sequence 25 AA;

Query Match 94.7%; Score 90; DB 23; Length 25;  
 Best Local Similarity 85.0%; Pred. No. 1.3e-08;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYKXHXSHQAHQVSLNS 20  
 ||||| | ||||| |||||  
 DB 1 IVGGYECTRHSHQAHQVSLNS 20  
 ||||| | ||||| |||||

RESULT 10  
 AAY93946  
 ID AAY93946 standard; peptide; 37 AA.  
 XX  
 AC AAY93946;  
 XX  
 DT 03-OCT-2000 (first entry)

RESULT 9  
 AAO21359  
 ID AAO21359 standard; Peptide; 25 AA.  
 AC AAO21359;  
 XX  
 XX 05-AUG-2002 (first entry)  
 XX Atlantic cod II multifunctional hydrolase peptide region.  
 KW Tranquilliser; vulnary; cytostatic; gynaecological; antiinflammatory;  
 KW corneal wound; internal trauma; polyanionic polymer; corneal ulcer;  
 KW adhesion; cervical spondylosis; organ membrane trauma; tissue; cavity;  
 KW endometriosis; pelvic inflammatory disease; cumulative trauma disorder; appendicitis;  
 KW peritonitis; pericarditis; pleuritis; inflammatory disease; adhesive peritonitis;  
 KW multifunctional protein; hydrolase; enzyme; atlantic cod.  
 OS Gadus morhua.  
 XX  
 PN WO200215913-A1.  
 XX  
 PD 28-FEB-2002.  
 XX  
 PF 23-AUG-2000; 2000WO-US23072.  
 XX  
 PR 23-AUG-2000; 2000WO-US23072.  
 XX  
 PA (PHAT-) PHAIRSON MEDICAL INC.  
 PA (EIDG-) EIDGENOSSISCHE TECH HOCHSCHULE ZUERICH.  
 XX  
 PI Franklin R, Cowling DSP, Hubbell JA, Van De Wetering P;  
 XX  
 DR WPI; 2002-425632/45.  
 XX  
 PT Treatment of corneal wounds, internal trauma or inflammatory diseases  
 PT comprises administration of a polyanionic polymer -  
 XX  
 PS Disclosure; Page 24; 59pp; English.  
 XX  
 CC The invention relates to a method for treating corneal wounds or internal  
 CC trauma. The method provides administration of a composition comprising a  
 CC polyanionic polymer. The method is useful for treating corneal wounds,  
 CC including corneal ulcers, corneal abrasions or injuries which may give  
 CC rise to corneal ulcers. The method is also useful for treating internal  
 CC trauma, including surgical wounds, trauma to a membrane that covers an  
 CC organ, tissue or cavity or trauma likely to give rise to adhesions. A  
 CC further method is useful for treating cervical spondylosis, cumulative  
 CC trauma disorder, endometriosis, pelvic inflammatory disease, adhesive  
 CC peritonitis, appendicitis, pericarditis, pleuritis, pericarditis and pleuritis or  
 CC other inflammatory diseases. This sequence represents the atlantic cod II  
 CC multifunctional hydrolase peptide region relating to the polyanionic  
 CC polymers of the invention.  
 XX  
 SQ Sequence 25 AA;

Query Match 94.7%; Score 90; DB 23; Length 25;  
 Best Local Similarity 85.0%; Pred. No. 1.3e-08;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYKXHXSHQAHQVSLNS 20  
 ||||| | ||||| |||||  
 DB 1 IVGGYECTRHSHQAHQVSLNS 20  
 ||||| | ||||| |||||

RESULT 10  
 AAY93946  
 ID AAY93946 standard; peptide; 37 AA.  
 XX  
 AC AAY93946;  
 XX  
 DT 03-OCT-2000 (first entry)

Wed Feb 12 11:59:30 2003

XX N-terminal of an atlantic cod trypsin hydrolase enzyme.  
 XX Transplantation rejection; hydrolase; graft versus host disease;  
 KW cell surface adhesion molecule; immune reaction; inflammation; shock;  
 KW tumour metastasis; autoimmune disease;  
 KW Krill derived multifunctional enzyme.  
 XX Gadus sp.  
 OS Key Location/Qualifiers  
 FH Misc-difference 5 /note= "any amino acid"  
 FT  
 FT  
 XX WO200038708-A1.  
 PN  
 XX  
 PD 06-JUL-2000.  
 XX  
 PF 23-DEC-1999; 99WO-US30818.  
 XX  
 PR 24-DEC-1998; 98US-0114147.  
 XX  
 XX (PHAI-) PHAIRSON MEDICAL INC.  
 PA Franklin RL, St Pierre Y;  
 PI  
 XX WPI; 2000-452301/39.  
 DR  
 XX Preventing or ameliorating transplantation rejection reactions using  
 XX hydrolase enzymes -  
 PT  
 XX Disclosure; Page 26; 66pp; English.  
 PS  
 XX The specification describes a method for preventing or ameliorating  
 CC transplantation rejection reactions for transplantation of immune cells  
 CC or other tissues. The method comprises treating a source of immune cells  
 CC with a hydrolase or hydrolase mixture and administering the treated  
 CC immune cells to a recipient animal. The hydrolase especially has a  
 CC preference for removing, destroying, inactivating or disabling at least  
 CC one of CD4, CD8, CD25, CD28, ICAM-1 (CD54), CD152, an integrin, CD154,  
 CC CD40 and CD80 in contrast to removing, destroying, inactivating or  
 CC disabling TCR. The methods are useful for preventing graft versus host  
 CC disease by using hydrolase enzymes to remove the cell surface adhesion  
 CC molecules which are involved in triggering the immune reactions involved  
 CC in the diseases. The methods are used for treating or preventing  
 CC cell-cell or cell-virus adhesion syndrome comprising inflammation, shock,  
 CC tumour metastases, autoimmune disease, transplantation rejection  
 CC reactions or microbial infections. The present sequence represents the  
 CC N-terminal of a trypsin hydrolase, which may be used in the method of  
 CC the invention.  
 XX  
 SQ Sequence 37 AA;  
 Query Match 94.7%; Score 90; DB 21; Length 37;  
 Best Local Similarity 85.0%; Pred. No. 2e-08; 0; Gaps 0;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 IVGGYXCXHSQAHOVSLSN 20  
 DB 1 IVGGYQCEAHSQAHOVSLSN 20  
 RESULT 11  
 AAO21360  
 ID AAO21360 standard; Peptide; 37 AA.  
 XX  
 AC AAO21360;  
 XX  
 XX 05-AUG-2002 (first entry)  
 DT  
 DE Atlantic cod multifunctional hydrolase N-terminal peptide region.  
 XX Tranquilliser; vulnery; cytostatic; gynaecological; antiinflammatory;  
 KW

KW corneal wound; internal trauma; polyanionic polymer; corneal ulcer;  
 KW corneal abrasion; surgical wound; organ membrane trauma; tissue; cavity;  
 KW adhesion; cervical spondylosis; cumulative trauma disorder; appendicitis;  
 KW endometriosis; pelvic inflammatory disease; adhesive peritonitis;  
 KW peritonitis; pericarditis; pleuritis; inflammatory peritonitis;  
 KW multifunctional protein; hydrolase; enzyme; atlantic cod.  
 XX  
 OS Gadus morhua.  
 XX  
 PN WO200215913-A1.  
 XX  
 PD 28-FEB-2002.  
 XX  
 PF 23-AUG-2000; 2000WO-US23072.  
 XX  
 PR 23-AUG-2000; 2000WO-US23072.  
 XX  
 XX (PHAI-) PHAIRSON MEDICAL INC.  
 PA (EIDG-) EIDGENOESSISCHE TECH HOCHSCHULE ZUERICH.  
 XX  
 PI Franklin R, Cowling DSP, Hubbell JA, Van De Wetering P;  
 XX  
 XX WPI; 2002-425632/45.  
 DR  
 XX Treatment of corneal wounds, internal trauma or inflammatory diseases  
 PT comprises administration of a polyanionic polymer -  
 PT  
 XX Disclosure; Page 24; 59pp; English.  
 PS  
 XX The invention relates to a method for treating corneal wounds or internal  
 CC trauma. The method provides administration of a composition comprising a  
 CC polyanionic polymer. The method is useful for treating corneal wounds,  
 CC including corneal ulcers, corneal abrasions or injuries which may give  
 CC rise to corneal ulcers. The method is also useful for treating internal  
 CC trauma, including surgical wounds, trauma to a membrane that covers an  
 CC organ, tissue or cavity or trauma likely to give rise to adhesions. A  
 CC further method is useful for treating cervical spondylosis, cumulative  
 CC trauma disorder, endometriosis, pelvic inflammatory disease, adhesive  
 CC peritonitis, appendicitis, pericarditis, pericarditis and pleuritis or  
 CC other inflammatory diseases. This sequence represents an N-terminal  
 CC peptide region of an atlantic cod multifunctional hydrolase relating to  
 CC the polyanionic polymers of the invention.  
 XX  
 SQ Sequence 37 AA;  
 Query Match 94.7%; Score 90; DB 23; Length 37;  
 Best Local Similarity 85.0%; Pred. No. 2e-08; 0; Gaps 0;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 IVGGYXCXHSQAHOVSLSN 20  
 DB 1 IVGGYQCEAHSQAHOVSLSN 20  
 RESULT 12  
 AAB31579  
 ID AAB31579 standard; peptide; 225 AA.  
 XX  
 AC AAB31579;  
 XX  
 XX 20-APR-2001 (first entry)  
 DT  
 XX Amino acid sequence of cod trypsin isozymes.  
 DE  
 XX Fish; serine proteinase; pain; acute osteoarthritis; chronic inflammation;  
 KW arthritis; inflamed joint; bursitis; osteoarthritis; septic arthritis;  
 KW rheumatoid arthritis; juvenile rheumatoid arthritis; fibromyalgia;  
 KW systemic lupus erythematosus; phlebitis; tendinitis; rash; psoriasis;  
 KW acne; eczema; facial seboreic eczema; foreskin infection;  
 KW athlete's foot; fistulae infection; ulcer; navel infection; wrinkle;  
 KW scar; keloid; boil; wart; allergic itch; hemorrhoid; wound;  
 KW fungal infection; autoimmune disease.  
 XX



Wed Feb 12 11:59:30 2003

protease activity; chymotrypsin; trypsin; collagenase; elastase; wound healing; corneal ulcer; internal trauma; internal surgical wound; nerve; tendon; sheath; adhesion formation inhibition; ophthalmological; vulnary.

Oncothynchus sp.

WO200049991-A2.

31-AUG-2000.

29-JUN-1999; 99WO-US14751.

23-FEB-1999; 99US-0256484.

(PHAI-) PHAIRSON MEDICAL INC.

Franklin R, Cowling D, Hubbell JA;

WPI; 2000-587120/55.

Use of microgel comprising crosslinked polyanionic polymer and optionally proteolytic enzyme in treatment of corneal wounds e.g. ulcers and abrasions and internal trauma e.g. surgical wounds and for treating implants to reduce adhesions -

Disclosure; Page 25; 85pp; English.

The invention relates to use of a microgel comprising a crosslinked polyanionic polymer in compositions for treatment of an area affected by a trauma selected from corneal wounds and internal trauma. The compositions further comprise a multifunctional protease having activities selected from two of chymotrypsin activity, trypsin activity, collagenase activity and elastase activity. In particular, the protease may be one of six isoforms of white shrimp (Penaeus vannamei) multifunctional hydrolase (AAB22947-B22952), or a multifunctional enzyme comprising one of the peptide sequences given in AAB22953-B22967. Compositions of the invention may be used to treat corneal ulcers (including infected ulcers), or abrasions or a chemical or physical insult to the cornea that are likely to give rise to a corneal ulcer. They are also used to treat internal traumas selected from an internal surgical wound, or a trauma to a membrane that covers either an internal organ or tissue or the cavity in which one or more internal organs or tissues reside, with the membrane being selected from the peritoneum, the pericardium, the epicardium and the pleura and a meninges. The internal trauma is to a tendon, tendon sheath, a nerve or a nerve sheath, where the trauma is susceptible to giving rise to adhesions. The amount of microgel administered is sufficient to prevent or reduce formation or reformation of adhesions, and the composition can also be used to treat an implant to reduce formation of adhesions. Sequences AAB22953-B22967 represent peptide sequences from a variety of multifunctional hydrolases. Multifunctional hydrolases comprising one of these sequences are claimed for use in compositions of the invention.

Query Match 82.1%; Score 78; DB 21; Length 25;  
Best Local Similarity 75.0%; Pred. No. 1.6e-06;  
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYKXCHSHQAHQVSLNS 20  
| | | | | : | | | | |  
Db 1 IVGGYCKAYSQAYQVSLNS 20

Search completed: February 12, 2003, 10:22:20  
Job time : 14.4328 secs

protease activity; chymotrypsin; trypsin; collagenase; elastase; wound healing; corneal ulcer; internal trauma; internal surgical wound; nerve; tendon; sheath; adhesion formation inhibition; ophthalmological; vulnary.

Oncothynchus sp.

WO200049991-A2.

31-AUG-2000.

29-JUN-1999; 99WO-US14751.

23-FEB-1999; 99US-0256484.

(PHAI-) PHAIRSON MEDICAL INC.

Franklin R, Cowling D, Hubbell JA;

WPI; 2000-587120/55.

Use of microgel comprising crosslinked polyanionic polymer and optionally proteolytic enzyme in treatment of corneal wounds e.g. ulcers and abrasions and internal trauma e.g. surgical wounds and for treating implants to reduce adhesions -

Disclosure; Page 25; 85pp; English.

The invention relates to use of a microgel comprising a crosslinked polyanionic polymer in compositions for treatment of an area affected by a trauma selected from corneal wounds and internal trauma. The compositions further comprise a multifunctional protease having activities selected from two of chymotrypsin activity, trypsin activity, collagenase activity and elastase activity. In particular, the protease may be one of six isoforms of white shrimp (Penaeus vannamei) multifunctional hydrolase (AAB22947-B22952), or a multifunctional enzyme comprising one of the peptide sequences given in AAB22953-B22967. Compositions of the invention may be used to treat corneal ulcers (including infected ulcers), or abrasions or a chemical or physical insult to the cornea that are likely to give rise to a corneal ulcer. They are also used to treat internal traumas selected from an internal surgical wound, or a trauma to a membrane that covers either an internal organ or tissue or the cavity in which one or more internal organs or tissues reside, with the membrane being selected from the peritoneum, the pericardium, the epicardium and the pleura and a meninges. The internal trauma is to a tendon, tendon sheath, a nerve or a nerve sheath, where the trauma is susceptible to giving rise to adhesions. The amount of microgel administered is sufficient to prevent or reduce formation or reformation of adhesions, and the composition can also be used to treat an implant to reduce formation of adhesions. Sequences AAB22953-B22967 represent peptide sequences from a variety of multifunctional hydrolases. Multifunctional hydrolases comprising one of these sequences are claimed for use in compositions of the invention.

Query Match 82.1%; Score 78; DB 20; Length 25;  
Best Local Similarity 75.0%; Pred. No. 1.6e-06;  
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 IVGGYKXCHSHQAHQVSLNS 20  
| | | | | : | | | | |  
Db 1 IVGGYCKAYSQAYQVSLNS 20

Search completed: February 12, 2003, 10:22:20  
Job time : 14.4328 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:22:36 ; Search time 34.2537 Seconds  
(without alignments)  
167.821 Million cell updates/sec

Title: US-10-036-371-8

Perfect score: 1158

Sequence: 1 IVGGVECTHSGAHQVSLNS.....GVYAKVYLSGVRDTRWVY 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 2554876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW PUB.pap.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW PUB.pap.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW PUB.pap.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pap.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW PUB.pap.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pap.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pap.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pap.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW PUB.pap.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pap.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW PUB.pap.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pap.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW PUB.pap.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1130	97.6	225	12	US-10-036-371-8
2	823.5	71.1	271	12	US-10-021-368-10
3	710	61.3	247	10	US-09-923-779-154
4	702	60.6	223	10	US-09-910-071-14
5	661.5	57.1	281	12	US-10-021-368-7
6	619.5	53.5	299	12	US-10-021-368-8
7	549.5	47.5	286	12	US-10-021-368-9
8	494	42.7	320	10	US-09-888-615-90
9	451	38.9	250	9	US-10-121-049-506
10	451	38.9	250	9	US-10-121-049-506
11	451	38.9	250	9	US-10-123-904-506
12	451	38.9	250	9	US-10-140-470-506
13	451	38.9	250	9	US-10-175-746-506
14	451	38.9	250	9	US-10-176-918-506
15	451	38.9	250	9	US-10-176-921-506
16	451	38.9	250	9	US-10-227-884-186
17	445	38.4	244	10	US-09-796-294-11
18	442	38.2	146	10	US-09-796-294-3
19	431	37.2	253	10	US-09-888-615-98

20	431	37.2	253	10	US-09-764-762-3
21	428.5	37.0	293	9	US-09-992-598-309
22	428.5	37.0	293	9	US-09-989-293A-309
23	428.5	37.0	293	9	US-09-989-735-309
24	428.5	37.0	293	9	US-09-990-444-309
25	428.5	37.0	293	9	US-09-989-730-309
26	428.5	37.0	293	9	US-09-990-436-309
27	428.5	37.0	293	9	US-09-991-181-309
28	428.5	37.0	293	9	US-09-993-687-309
29	428.5	37.0	293	9	US-09-989-734-309
30	428.5	37.0	293	9	US-10-028-072-456
31	428.5	37.0	293	9	US-09-997-653-309
32	428.5	37.0	293	9	US-09-933-667-309
33	428.5	37.0	293	9	US-10-121-049-456
34	428.5	37.0	293	9	US-10-123-904-456
35	428.5	37.0	293	9	US-10-140-470-456
36	428.5	37.0	293	9	US-09-990-438-309
37	428.5	37.0	293	9	US-09-997-428-309
38	428.5	37.0	293	9	US-09-990-562-309
39	428.5	37.0	293	9	US-09-997-428-309
40	428.5	37.0	293	9	US-10-175-746-456
41	428.5	37.0	293	9	US-10-176-918-456
42	428.5	37.0	293	9	US-10-176-921-456
43	428.5	37.0	293	10	US-09-739-907-82
44	428.5	37.0	293	10	US-09-739-907-97
45	428.5	37.0	293	10	US-09-989-722-309

#### ALIGNMENTS

RESULT 1  
US-10-036-371-8  
Sequence 8, Application US/10036371  
Patent No. US20020141987A1  
GENERAL INFORMATION:  
APPLICANT: BUARNARSON, JON B.  
TITLE OF INVENTION: FISH SERINE PROTEINASES AND THEIR PHARMACEUTICAL AND  
FILE REFERENCE: 81691/284960  
CURRENT APPLICATION NUMBER: US/10/036,371  
PRIOR FILING DATE: 1999-06-18  
PRIOR APPLICATION NUMBER: 09/411,688  
PRIOR FILING DATE: 1999-06-18  
PRIOR APPLICATION NUMBER: 5086/99  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 8  
LENGTH: 225  
TYPE: PRT  
ORGANISM: Gadus sp.  
FEATURE:  
NAME/KEY: MOD RES  
LOCATION: (9)  
OTHER INFORMATION: K or R  
FEATURE:  
NAME/KEY: MOD RES  
LOCATION: (24)  
OTHER INFORMATION: Y or F  
FEATURE:  
NAME/KEY: MOD RES  
LOCATION: (32)  
OTHER INFORMATION: K or E  
FEATURE:  
NAME/KEY: MOD RES  
LOCATION: (33)  
OTHER INFORMATION: D or Q  
FEATURE:  
NAME/KEY: MOD RES  
LOCATION: (64)  
OTHER INFORMATION: Y or F  
FEATURE:





Patent No. US20020106367A1  
GENERAL INFORMATION:  
APPLICANT: Band, Vimla  
TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
MOLECULES AND METHODS  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/021.368  
FILING DATE: 12-Dec-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/201.038  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00398/100002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 299 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-10-021-368-8  
Query Match 53.5%; Score 619.5; DB 12; Length 299;  
Best Local Similarity 50.2%; Pred. No. 1.1e-53;  
Matches 132; Conservative 32; Mismatches 56; Indels 43; Gaps 8;  
QY 1 IVGGYECTXH-----SQAHQVSLNSGVHXCGLSLNXXVWVSAACVYKSVLRVRLGEH 53  
DB 37 IVGGYHSTRYIVBVCENSLPYQVSLNSGSHFCGGLSISQWVWVSAACVYKTRIQVRLGEH 96  
QY 54 H-----IRVNEGTEQXISSSVYRHPNYSSYNI XNDIMLIKLTXPATL-----N 97  
DB 97 NHSTRYIVIKVLEGNQFINAKIIRHPKYRDTLDNDIMLIKUSSPAVHSTRYIVEN 156  
QY 98 QYVHAVALPTECAADATMCTVSGWNTMSVXD-GPKLQXLPILSHA-----DCA 148  
DB 157 ARVSTTSLTPAAGTECLISGWNTLSFGADYPPDELCLDAPVLTAHSTRYIVBECK 216  
QY 149 NSYGPGMITQSMFCAGYLEGGKSCQDGGPVVCGVGVVSWGY-----GCA 200  
DB 217 ASY-PGKITNSMFCVGFLEGGKSCQDGGPVVCGVGVVSWGY-----GCA 273  
QY 201 ERDXPGVYAKVXVLSGVRDTMA 223  
DB 274 WKNRPGVYTKVYNYVDWKDTIA 296  
RESULT 7  
US-10-021-368-9  
Sequence 9, Application US/10021368  
Patent No. US20020106367A1  
GENERAL INFORMATION:  
APPLICANT: Band, Vimla  
TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
MOLECULES AND METHODS  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/021.368  
FILING DATE: 12-Dec-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/201.038  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00398/100002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 286 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 9:  
US-10-021-368-9  
Query Match 47.5%; Score 549.5; DB 12; Length 286;  
Best Local Similarity 46.0%; Pred. No. 7.8e-47;  
Matches 121; Conservative 22; Mismatches 77; Indels 43; Gaps 7;  
QY 1 IVGGYE-----CTXHQAQVSLNSGVHXCGLSLNXXVWVSAACVYKSVLRVRLGE 52  
DB 25 IVGGYRTRYPVETCQEHSPYQVSLNAGSHICGGLITDQWVLSAACHVHPQVRLGE 84  
QY 53 HH-----IRVNEGTEQXISSSVYRHPNYSSYNI XNDIMLIKLTXPATL----- 96  
DB 85 HNRTRYPVBIYEIAGAQFIDAAKMLHPDYDKWTVDNDIMLIKUSSPATLNTRYPV 144  
QY 97 QYVHAVALPTECAADATMCTVSGWNTMSVXD-GPKLQXLPILSHAD-----CA 148  
DB 145 NSKVSTTLPQYCFATGTECLVSGWVGLKGFESPSVLQCLDAPVLSDSRNTRYPV 204  
QY 149 NSYGPGMITQSMFCAGYLEGGKSCQDGGPVVCGVGVVSWGY-----YGA 200  
DB 205 KAY-PRQITNNMFCVGFLEGGKSCQYDGGPVVCGVGVVSWGY-----GIVS 261  
QY 201 ERDXPGVYAKVXVLSGVRDTMA 223  
DB 262 LEGKPGVYTRVCNYLNWIQTVA 284  
RESULT 8  
US-09-888-615-90  
Sequence 90, Application US/09888615  
Patent No. US20020064856A1  
GENERAL INFORMATION:  
APPLICANT: PLOWMAN, GREGORY  
APPLICANT: WHYTE, DAVID



RESIST 9

```

RESULT 10
US/10-121-049-506
; Sequence 506, Application US/10121049
; Publication No. US2003002239A1
;
; GENERAL INFORMATION:
;
; APPLICANT: Baker, Kevin P.
; APPLICANT: Baresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
;
; TITLE OF INVENTION: SECRETED AND TR
;
; FILE REFERENCE: P3330RLC17
; CURRENT APPLICATION NUMBER: US/10/12
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - Sec
; NUMBER OF SEQ ID NOS: 550

```



APPLICANT: DeForge,Laura  
APPLICANT: Desnoyers,Luc  
APPLICANT: Filvaroff,Ellen  
APPLICANT: Gao,Wei-Qiang  
APPLICANT: Gerritsen,Mary E.  
APPLICANT: Goddard,Audrey  
APPLICANT: Godowski,Paul J.  
APPLICANT: Gurney,Austin L.  
APPLICANT: Sherwood,Steven  
APPLICANT: Smith,Victoria  
APPLICANT: Stewart,Timothy A.  
APPLICANT: Tumas,Daniel  
APPLICANT: Watanabe,Colin K  
APPLICANT: Wood,William  
APPLICANT: Zhang,Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3330R1C353  
CURRENT APPLICATION NUMBER: US/10/175,746  
CURRENT FILING DATE: 2002-06-19  
Prior Application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 550  
SEQ ID NO 506  
LENGTH: 250  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-175-746-506

Query Match 38.9%; Score 451; DB 9; Length 250;  
Best Local Similarity 41.0%; Pred. No. 3.3e-37;  
Matches 94; Conservative 38; Mismatches 87; Indels 10; Gaps 6;

QY 1 IVGGYECTXHSQAQVSL-NSGYHKGCSGLINXXVWVSAAHCKYKSVLRVRLGEHHIRVNE 59  
DB 22 IIKGFCEKPHSQPWAALFEKTRLLCGATLIAPRWLLTAHCKPKRYIVHLGQHNLOKEE 81  
QY 60 GTEQXISSSSVXRHPNYS---SYNIXNDIMLIKLTXPATLNQYVHAVALPTECAADATM 115  
DB 82 GCEQTRTATESPFGFNNSLPNKDHRNDIMLVKMASPVSIITWAVRPLTLSSRCVTAGTS 141  
QY 116 CTVSGWGNMTS-SVXDGDKLQKLXLPILSHADCANSGYSGMITQSMFCAGYLEGGKDSQC 174  
DB 142 CLISGWGSTSSPOLRPLHTLRCAITIEHOKCENAY-PGNITDTMVCASVOEGGKDSQC 200  
QY 175 GDSGGPVGCVNGVLOGVGVVSWGYG-CAERDXPGVYAKVXVLSGWRDYM 222  
DB 201 GDSGGPLVCNQSLO--GIISWGQDPCAITRKPGVYTKVCKYVDWIQETM 247

RESULT 14  
US-10-176-918-506  
Sequence 506, Application US/10176918  
Publication No. US20030027275A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: DeForge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3330R1C353  
CURRENT APPLICATION NUMBER: US/10/175,746  
CURRENT FILING DATE: 2002-06-19  
Prior Application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 550  
SEQ ID NO 506  
LENGTH: 250  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-175-746-506

FILE REFERENCE: P3330R1C382  
CURRENT APPLICATION NUMBER: US/10/176,918  
CURRENT FILING DATE: 2002-06-20  
Prior Application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 550  
SEQ ID NO 506  
LENGTH: 250  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-176-918-506

Query Match 38.9%; Score 451; DB 9; Length 250;  
Best Local Similarity 41.0%; Pred. No. 3.3e-37;  
Matches 94; Conservative 38; Mismatches 87; Indels 10; Gaps 6;

QY 1 IVGGYECTXHSQAQVSL-NSGYHKGCSGLINXXVWVSAAHCKYKSVLRVRLGEHHIRVNE 59  
DB 22 IIKGFCEKPHSQPWAALFEKTRLLCGATLIAPRWLLTAHCKPKRYIVHLGQHNLOKEE 81  
QY 60 GTEQXISSSSVXRHPNYS---SYNIXNDIMLIKLTXPATLNQYVHAVALPTECAADATM 115  
DB 82 GCEQTRTATESPFGFNNSLPNKDHRNDIMLVKMASPVSIITWAVRPLTLSSRCVTAGTS 141  
QY 116 CTVSGWGNMTS-SVXDGDKLQKLXLPILSHADCANSGYSGMITQSMFCAGYLEGGKDSQC 174  
DB 142 CLISGWGSTSSPOLRPLHTLRCAITIEHOKCENAY-PGNITDTMVCASVOEGGKDSQC 200  
QY 175 GDSGGPVGCVNGVLOGVGVVSWGYG-CAERDXPGVYAKVXVLSGWRDYM 222  
DB 201 GDSGGPLVCNQSLO--GIISWGQDPCAITRKPGVYTKVCKYVDWIQETM 247

RESULT 15  
US-10-176-921-506  
Sequence 506, Application US/10176921  
Publication No. US20030027276A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: DeForge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3330R1C288  
CURRENT APPLICATION NUMBER: US/10/176,921  
CURRENT FILING DATE: 2002-06-20  
Prior Application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 550  
SEQ ID NO 506  
LENGTH: 250  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-176-921-506

Query Match 38.9%; Score 451; DB 9; Length 250;  
Best Local Similarity 41.0%; Pred. No. 3.3e-37;  
Matches 94; Conservative 38; Mismatches 87; Indels 10; Gaps 6;

QY 1 IVGGYECTXHSQAQVSL-NSGYHKGCSGLINXXVWVSAAHCKYKSVLRVRLGEHHIRVNE 59

Db 22 I1KGECKPHSQPWAALFEKTRLLCGATLIAPRWLLTAACHLKPRYIVHVGHNLOKEE 81  
QY 60 GTEQXISSSVXRHPNYS----SYNIXNDIMLIKLTXPATLNQYVHVALPTECAADATM 115  
Db 82 GCEQTRTATESFPHPGFNNSLPNKDHENDIMLVKXASPVSIITWAVRPLTLSSRCVTAGTS 141  
QY 116 CTVSGWGNWMS-SVXDGDYLOXLPLILSHADCANSYFGMITOSMPCAGYLEGGKDSQC 174  
Db 142 CLISGWGSTSSPOLRPLHTLRCAINITIIIEHQKENAY-PGNITDTMVCASVOEGGKDSQC 200  
QY 175 GDSGGPVVNCVGLQGVVWSWYG-CAERDXPGYAKVXVLSGWRDYM 222  
Db 201 GDSGGPLVCNQLQ--GIISWGDPCAITRKEGVTKYKYVDWIGETM 247

Search completed: February 12, 2003, 10:31:08  
Job time : 35.2537 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:16:06 ; Search time 159.179 Seconds  
(without alignment)  
291.248 Million cell updates/sec

Title: US-10-036-371-8  
Perfect score: 1158  
Sequence: 1 IVGGVETXSHQAHQVLSLNS.....GVYAKVXVLSGWVDRNTMAXY 225

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21:.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	932.5	80.5	219	13 Q91036	Q91036 gadus morhu
2	915.5	79.1	242	13 Q9W7Q7	Q9W7Q7 paralicthy
3	910.5	78.6	242	13 Q93266	Q93266 pseudopleur
4	898.5	77.6	237	13 Q91515	Q91515 fugu rubrip
5	897.5	77.5	242	13 Q92099	Q92099 paranotothe
6	887	76.6	238	13 Q9W7Q6	Q9W7Q6 paralicthy
7	886	76.5	240	13 Q98TH0	Q98TH0 engraulis j
8	882	76.2	241	13 Q98TG9	Q98TG9 engraulis j
9	844.5	72.9	244	13 Q8QW3	Q8QW3 anguilla ja
10	740.5	63.9	244	13 Q42159	Q42159 petromyzon
11	737.5	63.7	247	13 Q42608	Q42608 petromyzon
12	736.5	63.6	245	13 Q42160	Q42160 petromyzon
13	736.5	63.6	247	13 Q42158	Q42158 petromyzon
14	736	63.6	246	11 Q9QUK9	Q9QUK9 mus musculu
15	735	63.5	246	11 Q9ROT7	Q9ROT7 mus musculu
16	733.5	63.3	178	13 Q93594	Q93594 dicentrarch

17	729	63.0	246	11 Q921R9	Q921R9 mus musculu
18	715	61.7	247	11 Q9CFN9	Q9CFN9 mus musculu
19	711	61.4	247	11 Q9D7Y7	Q9D7Y7 mus musculu
20	659.5	57.0	247	11 Q9CPN7	Q9CPN7 mus musculu
21	640	55.3	247	13 Q9W7Q5	Q9W7Q5 paralicthy
22	587	50.7	250	13 Q93265	Q93265 pseudopleur
23	581.5	50.2	249	13 Q9W6K0	Q9W6K0 notosthenia
24	563.5	48.7	249	13 Q9W6J8	Q9W6J8 dissostichu
25	560.5	48.4	249	13 Q92046	Q92046 dissostichu
26	548.5	47.4	344	13 Q9W6J9	Q9W6J9 dissostichu
27	498	43.0	264	5 Q62561	Q62561 penaeus van
28	493	42.6	255	4 Q96RQ0	Q96RQ0 homo sapien
29	488	42.1	263	5 Q62562	Q62562 penaeus van
30	474	40.9	263	5 Q9TX16	Q9TX16 penaeus van
31	474	40.9	266	5 Q27761	Q27761 penaeus van
32	471.5	40.7	267	5 Q9BK47	Q9BK47 luidia foli
33	471	40.7	266	5 Q6WR10	Q6WR10 paralthode
34	467.5	40.4	274	5 Q16133	Q16133 anopheles s
35	467.5	40.4	293	11 Q9D140	Q9D140 mus musculu
36	461.5	39.9	235	11 Q63274	Q63274 rattus norv
37	458.5	39.6	274	5 Q17086	Q17086 anopheles s
38	450	38.9	267	5 Q9VLF5	Q9VLF5 drosophila
39	442.5	38.2	234	11 Q9CV76	Q9CV76 mus musculu
40	442	38.2	249	11 Q9QYN4	Q9QYN4 mus musculu
41	442	38.2	276	11 Q9QYN3	Q9QYN3 mus musculu
42	439	37.9	146	13 Q9DD81	Q9DD81 brachydanio
43	438.5	37.9	239	11 Q63275	Q63275 rattus norv
44	437	37.7	268	5 Q46151	Q46151 pacifastacu
45	436.5	37.7	264	5 Q02569	Q02569 culex quing

## ALIGNMENTS

RESULT 1  
Q91036 PRELIMINARY; PRT; 219 AA.  
ID Q91036  
AC Q91036  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Trypsinogen I (Fragment)  
OS Gadus morhua (Atlantic cod)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.  
OX NCBI\_TaxID=8049;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ong T.L., Armstrong R.F., McNamara P., Buckley L.J.;  
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
DR EMBL; U47819; AAB02196.1; -.  
DR HSSP; P00763; LDPO.  
DR MEROPS; S01.125; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser.protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease.  
FT NON TER 1  
SQ SEQUENCE 219 AA; 23525 MW; C96964EB49CED1DA CRC64;

Query Match 80.5%; Score 932.5; DB 13; Length 219;  
Best Local Similarity 87.8%; Pred. No. 1.1e-90;  
Matches 176; Conservative 5; Mismatches 17; Indels 3; Gaps 2;

QY 1 IVGGVETXSHQAHQVLSLNSGVYKXVLSGWVDRNTMAXY 60

3

RESULT 4		PRELIMINARY; PRT; 237 AA.	
Q91515	ID Q91515		
AC	Q91515;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Trypsinogen (fragment)		
OS	Fugu rubripes (Japanese pufferfish) (Takifugu rubripes)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;		

Query Match	78.6%;	Score	910.5;	DB 13;	Length	242;
Best Local Similarity	75.6%;	Pred. NO. 2.7e-88;				
Matches 170.	Conservative	15;	Mismatches	37;	Indels	3;
					Gaps	2;



```

OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang K., Gan L., Lee I., Roach J., Hood L.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; U25747; AAA75001.1; -.
DR HSSP; P35031; 1B1T.
DR MEROPS; S01.125; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser. protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT NON TER 1
SQ SEQUENCE 237 AA; 25726 MW; 30D2DBAAC39080C2 CRC64;

Query Match 77.6%; Score 898.5; DB 13; Length 237;
Best Local Similarity 76.0%; Pred. No. 4.9e-87;
Matches 171; Conservative 13; Mismatches 38; Indels 3; Gaps 2;

QY 1 IVGGVECTXHSQAHQVSLNSGYKCGSLNXXVWVSAACHYKSVLRVLGHEHIRVNEG 60
Db 16 IVGGVECTXHSQAHQVSLNSGYKCGSLNXXVWVSAACHYKSVLRVLGHEHIRVNEG 75
QY 61 TEQXISSSVXRHPNYSYNXNDIMLIKLTXPATLNOYVHAVALPTECAADATMCTVSG 120
Db 76 TEQFISSSVIRHPNYSYNIDNIMLIKLSKPAFLNQYQVPALPSSCAAAGTMCKVSG 135
QY 121 WGNMTSSVXDGDKLQXLPILSHADCANSGVPGMITQSMFCAGYLEGGKDCSCQDSGGP 180
Db 136 WGNMTSSVXDGDKLQXLPILSHADCANSGVPGMITQSMFCAGYLEGGKDCSCQDSGGP 194
QY 181 WCNGLQGVGVSGVSGYGCARDXPGVYAKVXVLSGWVRDTMAXY 225
Db 195 WCNGLQGVGVSGVSGYGCARDXPGVYAKVXVLSGWVRDTMAXY 237

RESULT 5
Q20299 ID Q20299 PRELIMINARY; PRT; 242 AA.
AC TISSUE=PYLORIC CAECA;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsin precursor (EC 3.4.21.4).
OS Paranthothenia nagellanica (Maori cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
OC Notothenioidei; Nototheniidae; Paranthothenia.
OX NCBI_TaxID=37005;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PYLORIC CAECA;
RX MEDLINE=97104330; PubMed=8949488;
RA Gantoot S., Rencier-Belrue F., Edwards D., Van Beeumen J., Gerday C.;
RT "Trypsin and trypsinogen from an antarctic fish: molecular basis of
RL cold adaptation.";
RU Biochim. Biophys. Acta 1298:45-57 (1996).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; X82223; CAAS7701.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.125; -.
DR InterPro; IPR001314; Chymotrypsin.

```

```

DR InterPro; IPR001254; Ser. protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Signal.
FT SIGNAL 1 13 POTENTIAL.
SQ SEQUENCE 242 AA; 26201 MW; 3F4DE7CE80C4477C CRC64;

Query Match 77.5%; Score 897.5; DB 13; Length 242;
Best Local Similarity 75.1%; Pred. No. 6.4e-87;
Matches 169; Conservative 16; Mismatches 37; Indels 3; Gaps 2;

QY 1 IVGGVECTXHSQAHQVSLNSGYKCGSLNXXVWVSAACHYKSVLRVLGHEHIRVNEG 60
Db 21 IVGGVECTXHSQAHQVSLNSGYKCGSLNXXVWVSAACHYKSVLRVLGHEHIRVNEG 80
QY 61 TEQXISSSVXRHPNYSYNXNDIMLIKLTXPATLNOYVHAVALPTECAADATMCTVSG 120
Db 81 TEQFISSSVIRHPNYSYNIDNIMLIKLSKPAFLNQYQVPALPSSCAAAGTMCTVSG 140
QY 121 WGNMTSSVXDGDKLQXLPILSHADCANSGVPGMITQSMFCAGYLEGGKDCSCQDSGGP 180
Db 141 WGNMTSSVXDGDKLQXLPILSHADCANSGVPGMITQSMFCAGYLEGGKDCSCQDSGGP 199
QY 181 WCNGLQGVGVSGVSGYGCARDXPGVYAKVXVLSGWVRDTMAXY 225
Db 200 WCNGLQGVGVSGVSGYGCARDXPGVYAKVXVLSGWVRDTMAXY 242

RESULT 6
Q20706 ID Q20706 PRELIMINARY; PRT; 238 AA.
AC Q20706;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsinogen 2 (Fragment).
OS Paralichthys olivaceus (Flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoides; Paralichthyidae; Paralichthys.
OX NCBI_TaxID=8255;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RA Suzuki T., Srivastava A.S., Kurokawa T.;
RL "Japanese flounder mRNA for trypsinogen 2.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB029751; BAA82363.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.125; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser. protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT NON TER 1
SQ SEQUENCE 238 AA; 26071 MW; F2B8908085B8D062 CRC64;

Query Match 76.6%; Score 887; DB 13; Length 238;
Best Local Similarity 75.1%; Pred. No. 8.1e-86;
Matches 169; Conservative 14; Mismatches 38; Indels 4; Gaps 3;

```

Wed Feb 12 11:59:48 2003

us-10-036-371-8.rsp

```
QY 1 IVGGYECTHSQAQVSLNSGYHXCGLSLNXXVWVSAACHYKSVLRVRLGHEHHRVNEG 60
DB 18 IVGGYECTPVSPHQVSLNSGYHFCGSLVNWVWVSAACHYKSVLRVRLGHEHHRVNEG 77
QY 61 TEQXISSSSVKRPNPVSSNYXNDIMLIKLTXPATLNQVYHVALPTECAADATMCTVSG 120
DB 78 TEQFIDSSVRIRHPNPVSSNYXNDIMLIKLTSEPATLNQVYHVALPTECAPAGTMCTVSG 137
QY 121 WGMTSMNVSGDKLQKXLPILSHADCANSGPGMITQSMFCAGYLEGGKDSQCGDSGGP 180
DB 138 WGMTSMNV--TDSRLQCLDLPILSERDCNSY-PGMITNMFACAGYLEGGKDSQCGDSGGP 195
QY 181 VVNCNGVLQGVVSWGVCACERDHPGVYAKVXVLSGWRDRTMAXY 225
DB 196 VVNCNGVLQ--GVVSWGVCACERDHPGVYAKVCFIDWLTERTMSY 238

RESULT 7
Q98TH0 PRELIMINARY; PRT; 240 AA.
AC Q98TH0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsinogen.
OS Engraulis japonicus (Japanese anchovy).
OC Eukaryota; Metazoa; Chordata; Craniata;
OC Actinopterygii; Neopterygii; Teleostei; Clupeomorpha; Engraulidae;
OC Engraulis.
OX NCBI_TaxID=42892;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PYLORIC CAECA;
RA Watabe S., Ahsan M.N., Funabara D.;
RT "Anchovy trypsinogen mRNA.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB041929; BAB40329.1; -.
DR HSSP; P00763; 1DPO.
DR MEROPS; S01.125; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin_1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00134; TRYPSIN_DOM; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 240 AA; 26026 MW; 299B119BFF071464 CRC64;

Query Match 76.5%; Score 886; DB 13; Length 240;
Best Local Similarity 74.9%; Pred. No. 1e-85;
Matches 167; Conservative 14; Mismatches 38; Indels 4; Gaps 3;

QY 1 IVGGYECTHSQAQVSLNSGYHXCGLSLNXXVWVSAACHYKSVLRVRLGHEHHRVNEG 60
DB 20 IVGGYECQHPVSPHQVSLNSGYHFCGSLVNWVWVSAACHYKSVLRVRLGHEHHRVNEG 79
QY 61 TEQXISSSSVKRPNPVSSNYXNDIMLIKLTXPATLNQVYHVALPTECAADATMCTVSG 120
DB 80 TEQFIDSSVRIRHPNPVSSNYXNDIMLIKLTSEPATLNQVYHVALPTECAPAGTMCTVSG 139
QY 121 WGMTSMNVSGDKLQKXLPILSHADCANSGPGMITQSMFCAGYLEGGKDSQCGDSGGP 180
DB 140 WGMTSMNV--SGDKLQCLQIPILSDRDCNSY-PGMITDAMFCAGYLEGGKDSQCGDSGGP 197
QY 181 VVNCNGVLQGVVSWGVCACERDHPGVYAKVXVLSGWRDRTMAXY 223
DB 196 VVNCNGVLQ--GVVSWGVCACERDHPGVYAKVCFIDWLTERTMSY 238
```

```
RESULT 8
Q98TG9 PRELIMINARY; PRT; 241 AA.
AC Q98TG9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsinogen II.
GN ATR11.
OS Engraulis japonicus (Japanese anchovy).
OC Eukaryota; Metazoa; Chordata; Craniata;
OC Actinopterygii; Neopterygii; Teleostei; Clupeomorpha; Engraulidae;
OC Engraulis.
OX NCBI_TaxID=42892;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PYLORIC CAECA;
RA Watabe S., Ahsan M.N., Funabara D.;
RT "Anchovy trypsinogen mRNA.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB041930; BAB40330.1; -.
DR HSSP; P00763; 1DPO.
DR MEROPS; S01.258; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin_1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00134; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease.
KW Hydrolase; Serine protease.
SQ SEQUENCE 241 AA; 26282 MW; FE362D39CAEBE2F6 CRC64;

Query Match 76.2%; Score 882; DB 13; Length 241;
Best Local Similarity 73.8%; Pred. No. 2.8e-85;
Matches 166; Conservative 15; Mismatches 40; Indels 4; Gaps 3;

QY 1 IVGGYECTHSQAQVSLNSGYHXCGLSLNXXVWVSAACHYKSVLRVRLGHEHHRVNEG 60
DB 20 IVGGYECQHPVSPHQVSLNSGYHFCGSLVSWVSAACHYKSVLRVRLGHEHHRVNEG 79
QY 61 TEQXISSSSVKRPNPVSSNYXNDIMLIKLTXPATLNQVYHVALPTECAADATMCTVSG 120
DB 80 NEQFIDSSVRIRHPNPVSSNYXNDIMLIKLTSEPATLNQVYHVALPTECAPAGTMCTVSG 139
QY 121 WGMTSMNVSGDKLQKXLPILSHADCANSGPGMITQSMFCAGYLEGGKDSQCGDSGGP 180
DB 140 WGMTSMNV--SGDKLQCLQIPILSDRDCNSY-PGMITESMFCAGYLEGGKDSQCGDSGGP 197
QY 181 VVNCNGVLQGVVSWGVCACERDHPGVYAKVXVLSGWRDRTMAXY 225
DB 198 VVNCNGVLQ--GVVSWGVCACERDHPGVYAKVCLFNDWIDSTMAQY 240

RESULT 9
Q98GW3 PRELIMINARY; PRT; 244 AA.
AC Q98GW3;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsinogen (EC 3.4.21.4).
GN TRY.
OS Anguilla japonica (Japanese eel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguillidae;
OC Anguilla.
OX NCBI_TaxID=7937;
RN [1]
RP SEQUENCE FROM N.A.
```



OC Petromyzontiformes; Petromyzontidae; Petromyzon.

ON NCBI\_TaxID=7757;

OX [1]

RY SEQUENCE FROM N.A.

RA Roach J.C.;

RL "The Molecular Evolution of the Vertebrate Trypsinogens";

RT Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

CC TRYPSIN FAMILY.

CC EMBL; AF011898; AAB69654.1; -.

DR HSSP; P00763; IDPO.

DR MEROPS; S01.128; -.

DR InterPro; IPR001314; Chymotrypsin.

DR InterPro; IPR001254; Ser.protease\_Try.

DR Pfam; PF00689; trypsin; 1.

DR PRINTS; P00072; CHYMOTRYPSIN.

DR SMART; SM00020; TRYPSIN\_DOM; 1.

DR PROSITE; PS00240; TRYPSIN\_DOM; 1.

DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.

DR PROSITE; PS00135; TRYPSIN\_SER; 1.

KW Hydrolase; Serine protease; Signal.

FT SIGNAL 1 35 POTENTIAL

FT CHAIN 16 247 TRYPSIN A2.

FT SEQUENCE 247 AA; 26309 MW; AD73E88531970324 CRC64;

Query Match 63.6%; Score 736.5; DB 13; Length 247;

Best Local Similarity 61.8%; Pred. No. 7.1e-70;

Matches 139; Conservative 22; Mismatch 59; Indels 5; Gaps

Qy 1 IVGGYETXHSQAHQVSLNSGYKXCGSLINXXVWVYSAACHYKSVLR--VRLGEHHRVN 58

Db 24 IVGGSECAHSQPVQVSLNTGYHFCGSLINSQVWVYSAACHYQTSRISVRIGEHNIFVN 83

Qy 59 EGTROXTSSSSVVRHPPVSYNNKNDIMLIKTKPATLQYHVAVALPTECAADATMCTV 118

Db 84 EGTGQQQASKAIQHPQYNWTNDIMLIKLSPTATLQYQAQIALPSSCVNTGVWCTI 143

Qy 119 SGWNTMSSVXGDKQLQXLPILSHADCANSYGPGMITQSMFCAGLEGKDCSCGDSG 178

Db 144 SGWGETSTIGSPDVLVCVQAPVLSDTSCNSY-PGDIITNNMTCGLEGKDCSCGDSG 202

Qy 179 GPVVCNGVLQGVVSGVNGYCAERDXPGVVAKVXVLSGWVRDTMA 223

Db 203 GPVVCNGELQ--GIVSWGRCALPNYGPVYTKVCYNNAWTAQTIA 245

RESULT 14

Q9QUK9 PRELIMINARY; PRT; 246 AA.

AC Q9QUK9;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE TESP4 (0910001B19RIK protein) (Trypsinogen 9).

GN TC OR 0910001B19RIK OR TRYPSINOGEN.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=TESTIS;

RX MEDLINE=99436155; PubMed=10506205;

RA Ohmura K., Kohno N., Kobayashi Y., Yamagata K., Sato S.,

RA Kashiwabara S., Baba T.;

RT "A homologue of pancreatic trypsin is localized in the acrosome of

RT mammalian sperm and is released during acrosome reaction.";

RL J. Biol. Chem. 274:29426-29432(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=PANCREAS;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,



Qy 1 IVGGYECTHSHQAHVSLNSGYHXCGLINXXVWVSAHCKSVLRLGEHHIRVNEG 60  
Db 24 IVGGYTCRENSVPYQVSLNSGYHFCGSLINDOMVWVSAHCKSVRIQVRLGEHNINVLEG 83  
Qy 61 TEOXISSSVKRPNTSSYNIXNDIMLIKLTXPATLNOYVHAVALPTECAADATMCTVSG 120  
Db 84 NEQFVNSAKIHKPNFNSRTLNNDIMLIKLASPTLNARVATVALPSSCAPACTQCLISG 143  
Qy 121 WGNMTS-SVXDGDKLQXLPLILSHADCANSYGFPMITQSMFCAGYLEGGKDSQGDSDGG 179  
Db 144 WGNTLFSGVNPDLLOCLDAPLLPQADCEASY-PGKITNNMICVGFLEGGKDSQGDSDGG 202  
Qy 180 PVVCNGVLOGVGVSVWGYGCAERDXPGVYAKVXVLSGWRDTWA 223  
Db 203 PVVCNGQLQ--GIVSWGYGKALKNPGVYTKVCNVDWIQNTIA 244

Search completed: February 12, 2003, 10:27:30  
Job time : 160.179 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:18:40 ; Search time 48.3582 Seconds  
(without alignments)  
136.898 Million cell updates/sec

Title: US-10-036-371-8

Perfect score: 1158

Sequence: 1 IVGGYECTXHSQAHQVSLNS.....GVYAKVVLUSGMVDRITMAXY 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/6C\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/6D\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	823.5	71.1	271	1	US-08-467-155A-10
2	823.5	71.1	271	2	US-08-628-198-7
3	823.5	71.1	271	4	US-09-201-038-10
4	823.5	71.1	271	5	PCT-US96-07343-10
5	755	65.2	246	2	US-08-978-404B-44
6	710	61.3	224	4	US-08-944-483-34
7	707	61.1	247	2	US-08-956-267A-2
8	702	60.6	224	4	US-08-944-483-36
9	702	60.6	224	4	US-08-766-982-13
10	702	60.6	224	4	US-09-296-219-13
11	702	60.6	225	2	US-09-027-337-5
12	702	60.6	225	4	US-09-644-600-5
13	701	60.5	223	1	US-08-278-091-9
14	701	60.5	223	1	US-08-483-859-9
15	701	60.5	223	1	US-08-472-173-9
16	701	60.5	223	2	US-08-487-167-9
17	701	60.5	223	2	US-08-482-816-9
18	701	60.5	223	2	US-08-296-149-9
19	701	60.5	223	2	US-08-801-499-9
20	701	60.5	223	2	US-08-615-271-9
21	701	60.5	223	3	US-09-074-660-9
22	701	60.5	223	3	US-09-074-659-9
23	701	60.5	223	3	US-09-106-468-9
24	701	60.5	223	4	US-09-106-466A-9
25	701	60.5	223	4	US-09-106-467-9
26	701	60.5	229	3	US-09-120-582-2
27	691	59.7	224	4	US-08-944-483-35

28	661.5	57.1	281	1	US-08-467-155A-7	Sequence 7, Appli
29	661.5	57.1	281	2	US-08-628-198-7	Sequence 7, Appli
30	661.5	57.1	281	4	US-09-201-038-7	Sequence 7, Appli
31	661.5	57.1	281	5	PCT-US96-07343-7	Sequence 7, Appli
32	619.5	53.5	299	1	US-08-467-155A-8	Sequence 8, Appli
33	619.5	53.5	299	2	US-08-628-198-8	Sequence 8, Appli
34	619.5	53.5	299	4	US-09-201-038-8	Sequence 8, Appli
35	619.5	53.5	299	5	PCT-US96-07343-8	Sequence 8, Appli
36	549.5	47.5	286	1	US-08-467-155A-9	Sequence 9, Appli
37	549.5	47.5	286	2	US-08-628-198-9	Sequence 9, Appli
38	549.5	47.5	286	4	US-09-201-038-9	Sequence 9, Appli
39	549.5	47.5	286	5	PCT-US96-07343-9	Sequence 9, Appli
40	486	42.0	156	4	US-09-261-416-6	Sequence 9, Appli
41	466	40.2	151	4	US-09-518-046-21	Sequence 21, Appli
42	451	38.9	282	3	US-09-025-059-1	Sequence 1, Appli
43	450	38.9	248	4	US-08-944-483-24	Sequence 24, Appli
44	449.5	38.8	289	4	US-09-386-642-14	Sequence 14, Appli
45	448	38.7	244	1	US-08-361-395-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-08-467-155A-10

; Sequence 10, Application US/08467155A

; Patent No. 5736377

; GENERAL INFORMATION:

; APPLICANT: Band, Vimla

; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/467,155A

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Clark, Paul T.

; REGISTRATION NUMBER: 30,162

; REFERENCE/DOCKET NUMBER: 00398/100001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617/542-5070

; TELEFAX: 617/542-8906

; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 271 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-467-155A-10

Query Match 71.1%; Score 823.5; DB 1; Length 271;  
Best Local Similarity 63.4%; Pred. No. 2,2e-86;  
Matches 168; Conservative 16; Mismatches 38; Indels 43; Gaps 7;

QY 1 IVGGY-----ECTXHSQAHQVSLNSGYHFCGSLVNNWVVAHCKYKSVLRVLGE 52

Db 10 IVGGYSTYPIIECKAYSQHVSLNSGYHFCGSLVNNWVVAHCKYKSVLRVLGE 69

Wed Feb 12 11:59:46 2003

QY 53 HH-----IRNEGTEQXISSSVKRRHPNYSYNIXNDIMLIKLTXPATL----- 96  
Db 70 HNSSTRPIIIQVTEGSEQFISSSRVIRHPNYSYNIDNDIMLIKLSKPATLSSTRYP11 129  
QY 97 NOYHVALPTECAADATMCTVSGWNTMSVVXGDKLOXLKLPILSHA-----DCA 148  
Db 130 NTVQVVALPTSCAPAGTMTCTVSGWNTMSSTADKNKLOCLNIPILSYSSSTRYP11DCN 189  
QY 149 NSYGPMTQSMFCAGYLEGKSCQSDSGPVVNCVGLQGVVSWGY-----GCA 200  
Db 190 NSY-PGMITNAMFCAGYLEGKSCQSDSGPVVNCVGLQ--GVWSGYSSSTRYP11GCA 246  
QY 201 ERDXPGVYAKVXVLSGWVDTMAXY 225  
Db 247 EPGNPGVYAKVCIENDWLTSTMATY 271

RESULT 2  
US-08-628-198-10  
; Sequence 10, Application US/08628198  
; Patent No. 5843694  
; GENERAL INFORMATION:  
; APPLICANT: Band, Vimla  
; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
; TITLE OF INVENTION: MOLECULES AND METHODS  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/628,198  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/467,155  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00398/100002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 271 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-628-198-10

Query Match 71.1%; Score 823.5; DB 2; Length 271;  
Best Local Similarity 63.4%; Pred. No. 2.2e-86;  
Matches 168; Conservative 16; Mismatches 38; Indels 43; Gaps 7;  
QY 1 IVGGY-----ECTXHSQAQVLSNGYHXCXGSLINXKXWVWSAAHCYKSVLRVRLGE 52  
Db 10 IVGGYSTRYP11IECKAYSQPHQVSLNSGYHPCGGSILNENWVWSAAHCYQSRVEVRLGE 69  
QY 53 HH-----IRNEGTEQXISSSVKRRHPNYSYNIXNDIMLIKLTXPATL----- 96  
Db 70 HNSSTRYP11IIVTEGSEQFISSSRVIRHPNYSYNIDNDIMLIKLSKPATLSSTRYP11 129

QY 97 NOYHVALPTECAADATMCTVSGWNTMSVVXGDKLOXLKLPILSHA-----DCA 148  
Db 130 NTVQVVALPTSCAPAGTMTCTVSGWNTMSSTADKNKLOCLNIPILSYSSSTRYP11DCN 189  
QY 149 NSYGPMTQSMFCAGYLEGKSCQSDSGPVVNCVGLQGVVSWGY-----GCA 200  
Db 190 NSY-PGMITNAMFCAGYLEGKSCQSDSGPVVNCVGLQ--GVWSGYSSSTRYP11GCA 246  
QY 201 ERDXPGVYAKVXVLSGWVDTMAXY 225  
Db 247 EPGNPGVYAKVCIENDWLTSTMATY 271

RESULT 3  
US-09-201-038-10  
; Sequence 10, Application US/09201038  
; Patent No. 6153387  
; GENERAL INFORMATION:  
; APPLICANT: Band, Vimla  
; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
; TITLE OF INVENTION: MOLECULES AND METHODS  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/201,038  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/628,198  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00398/100002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 271 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-201-038-10

Query Match 71.1%; Score 823.5; DB 4; Length 271;  
Best Local Similarity 63.4%; Pred. No. 2.2e-86;  
Matches 168; Conservative 16; Mismatches 38; Indels 43; Gaps 7;  
QY 1 IVGGY-----ECTXHSQAQVLSNGYHXCXGSLINXKXWVWSAAHCYKSVLRVRLGE 52  
Db 10 IVGGYSTRYP11IECKAYSQPHQVSLNSGYHPCGGSILNENWVWSAAHCYQSRVEVRLGE 69  
QY 53 HH-----IRNEGTEQXISSSVKRRHPNYSYNIXNDIMLIKLTXPATL----- 96  
Db 70 HNSSTRYP11IIVTEGSEQFISSSRVIRHPNYSYNIDNDIMLIKLSKPATLSSTRYP11 129  
QY 97 NOYHVALPTECAADATMCTVSGWNTMSVVXGDKLOXLKLPILSHA-----DCA 148



Db 130 NTYQVQVALPTSCAPAGTCTVSGWNTMSADTKNKLQCLNIPILSYSSSTRYPIIDCN 189  
 Qy 149 NSYGPGMITQSMFCAGYLEGGKDCQSGDGGPPVVCNGVLQGVVSWGY-----GCA 200  
 Db 190 NSY-PGMITNAMFCAGYLEGGKDCQSGDGGPPVVCNGELQ--GVVSWGYSTRYPIIGCA 246  
 Qy 201 ERDXPGVYAKVXVLSGWRDWTMAXY 225  
 Db 247 EPGNPGVYAKVCFNDWLTSTMATY 271

RESULT 4  
 PCT-US96-07343-10  
 ; Sequence 10, Application PC/TUS9607343  
 ; GENERAL INFORMATION:  
 ; APPLICANT: New England Medical Center Hospitals, Inc.  
 ; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
 ; TITLE OF INVENTION: MOLECULES AND METHODS  
 ; NUMBER OF SEQUENCES: 11  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.  
 ; STREET: 225 Franklin Street  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02110-2804  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US96/07343  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/467,155  
 ; FILING DATE: 06-JUN-1995  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Clark, Paul T.  
 ; REGISTRATION NUMBER: 30,162  
 ; REFERENCE/DOCKET NUMBER: 00398/100001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617/542-5070  
 ; TELEFAX: 617/542-8906  
 ; TELEX: 200154  
 ; INFORMATION FOR SEQ ID NO: 10:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 271 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; PCT-US96-07343-10

Query Match 71.1%; Score 823.5; DB 5; Length 271;  
 Best Local Similarity 63.4%; Pred. No. 2.2e-86;  
 Matches 168; Conservative 16; Mismatches 38; Indels 43; Gaps 7;  
 Qy 1 IVGGY-----ECTHSQAHQVLSNGYHXCGSLINXXWVWSAAHCYKSVLRVRLGE 52  
 Db 10 IVGGYSTRYPIIECKAYSQPHQVLSNGYHFCGSLVNVWVWSAAHCYQSRVEVRLGE 69  
 Qy 53 RH-----IRVNEGTEQKISSSVXSRHPNYSSYNIDIMLIKLTYPATL----- 96  
 Db 70 HNSSTRYPIIQVTEGSEQFISSRVIRHPNYSSYNIDIMLIKLTYPATLSTRYPII 129  
 Qy 97 NQVHVALPTSCAADATCTVSGWNTMSVXDGDKLQXKLPLILSHA-----DCA 148  
 Db 130 NTYQVQVALPTSCAPAGTCTVSGWNTMSADTKNKLQCLNIPILSYSSSTRYPIIDCN 189  
 Qy 149 NSYGPGMITQSMFCAGYLEGGKDCQSGDGGPPVVCNGVLQGVVSWGY-----GCA 200

Db 190 NSY-PGMITNAMFCAGYLEGGKDCQSGDGGPPVVCNGELQ--GVVSWGYSTRYPIIGCA 246  
 Qy 201 ERDXPGVYAKVXVLSGWRDWTMAXY 225  
 Db 247 EPGNPGVYAKVCFNDWLTSTMATY 271  
 RESULT 5  
 US-08-978-404B-44  
 ; Sequence 44, Application US/08978404B  
 ; Patent No. 5968782  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Stevens, Richard L.  
 ; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
 ; TITLE OF INVENTION: FIBRINOGEN  
 ; NUMBER OF SEQUENCES: 74  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
 ; STREET: 600 Atlantic Avenue  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: U.S.A.  
 ; ZIP: 02210-2211  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/978,404B  
 ; FILING DATE: 25-NOV-97  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/032,354  
 ; FILING DATE: 04-DEC-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Plumer, Elizabeth R.  
 ; REGISTRATION NUMBER: 36,637  
 ; REFERENCE/DOCKET NUMBER: B0801/7090  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-720-3500  
 ; TELEFAX: 617-720-2441  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 44:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 246 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: No. 5968782e  
 ; US-08-978-404B-44  
 Query Match 65.2%; Score 755; DB 2; Length 246;  
 Best Local Similarity 63.8%; Pred. No. 1.4e-78;  
 Matches 143; Conservative 25; Mismatches 52; Indels 4; Gaps 3;  
 Qy 1 IVGGYECTHSQAHQVLSNGYHXCGSLINXXWVWSAAHCYKSVLRVRLGEHHRVNEG 60  
 Db 24 IVGGYTCPEHSVPYQVLSNGYHFCGSLINQWVWSAAHCYKSRICQVRLGEHNVLEG 83  
 Qy 61 TEQKISSSVXSRHPNYSSYNIDIMLIKLTYPATLNOYHVALPTECAADATMCTVSG 120  
 Db 84 DEQFINAAKTIKHPNYSSWTLNNDIMLIKLSPPVKLNARVAPVALPSACAPAGTQCLISG 143  
 Qy 121 WGNMTSS-VXDGDKLQXKLPLILSHADCANSGPGMITQSMFCAGYLEGGKDCQSGDGG 179  
 Db 144 WGNLTSGVNVNPDLLQCVDAFVLSQADCEAY-PGEITSSMICVGFLEGGKDCQSGDGG 202  
 Qy 180 PWCNGLVQGVVSWGYGCAERDXPGVYAKVXVLSGWRDWTMA 223  
 Db 203 PWCNGLQ--GIVSWGYGCAFPDNPVYTKVCFNVFGMIQDTIA 244

RESULT 6  
US-08-944-483-34  
; Sequence 34, Application US/08944483  
; Patent No. 6232456  
; GENERAL INFORMATION:  
; APPLICANT: COHEN, MAURICE  
; APPLICANT: COUPITTS, TRACEY L.  
; APPLICANT: FRIEDMAN, PAULA N.  
; APPLICANT: GRANADOS, EDWARD N.  
; APPLICANT: KLASS, MICHAEL R.  
; APPLICANT: RUSSELL, JOHN C.  
; APPLICANT: STEWART, KENT D.  
; APPLICANT: STROUPE, STEVEN D.  
; TITLE OF INVENTION: NOVEL SRINE PROTEASE REAGENTS  
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES  
; TITLE OF INVENTION: OF THE PROSTATE  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/944,483  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Becker, Cheryl L.  
; REGISTRATION NUMBER: 35,441  
; REFERENCE/DOCKET NUMBER: 6183.US.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 847/935-1729  
; TELEFAX: 847/938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 224 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 6232456e  
US-08-944-483-34  
Query Match 61.3%; Score 710; DB 4; Length 224;  
Best Local Similarity 60.3%; Pred. No. 1.7e-73;  
Matches 135; Conservative 31; Mismatches 54; Indels 4; Gaps 3;  
Qy 1 IVGGYECTHSHQAHVSLNSGYHXCGLSLINXXWVSAACHYKSVLRVRLGHHIRVNEG 60  
Db 1 IVGGYTCENSPYQVSLNSGYHFCGSLISBQWVVSAGHCYKSRQVRLGHHIRVNEG 60  
Qy 61 TEQXISSSVXKHPNYSYINXNDIMLIKLTXPATLNQYVHAVALPTECAADATMCTVSG 120  
Db 61 NEQFINAKIIRPKNSRNLNDILLIKLSPAVINSRVSAISLPTAPPAAGTESLISG 120  
Qy 121 WNTMSSVXD-GDKLQXLPLILSHADCANSGYCGMITQSMFCAGVLEGGKSCQGDSSG 179  
Db 121 WNTLSSGADYPDELQCLDAPVLSQACEASY-PGKITNNMFCVGFLEGGKSCQGDSSG 179  
Qy 180 PVVNCVLOGVGVVSGYGAERDXPGVYAKVXVLSGWVDRDTMA 223  
Db 180 PVVNSGELQ--GIVSWGYGCAQKNRPGVYTKVNYVDWINDTIA 221

RESULT 7  
US-08-956-267A-2  
; Sequence 2, Application US/08956267A  
; Patent No. 5945328  
; GENERAL INFORMATION:  
; APPLICANT: WOLDIKE, Helle Fabricius  
; APPLICANT: KJELDSEN, Thomas Borglum  
; TITLE OF INVENTION: A Process For Producing Trypsin  
; TITLE OF INVENTION: (Trypsinogen)  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 59453280 No. 5945328disk of No. 5945328th America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/956,267A  
; FILING DATE: 22-OCT-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rozek, Carol  
; REGISTRATION NUMBER: 36,993  
; REFERENCE/DOCKET NUMBER: 4500.204-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 247 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-956-267A-2  
Query Match 61.1%; Score 707; DB 2; Length 247;  
Best Local Similarity 60.3%; Pred. No. 4.4e-73;  
Matches 135; Conservative 27; Mismatches 58; Indels 4; Gaps 3;  
Qy 1 IVGGYECTHSHQAHVSLNSGYHXCGLSLINXXWVSAACHYKSVLRVRLGHHIRVNEG 60  
Db 25 IVGGYTCAANSIPYQVSLNSGSHFCGSLINSQWVVSAAHCYKSRQVRLGHHIRVNEG 84  
Qy 61 TEQXISSSVXKHPNYSYINXNDIMLIKLTXPATLNQYVHAVALPTECAADATMCTVSG 120  
Db 85 NEQFINAKIIRPKNSRNLNDILLIKLSPATLNSRVATVSLPRSCAAAGTECLISG 144  
Qy 121 WNTMSSVXD-GDKLQXLPLILSHADCANSGYCGMITQSMFCAGVLEGGKSCQGDSSG 179  
Db 145 WNTTKSSGSSPYSLQCLKAPVLSDSCKSSY-PGQITGNMICVGFLEGGKSCQGDSSG 203  
Qy 180 PVVNCVLOGVGVVSGYGAERDXPGVYAKVXVLSGWVDRDTMA 223  
Db 204 PVVNCGLQ--GIVSWGYGCAQKNRPGVYTKVNYVDWINDTIA 245  
RESULT 8  
US-08-766-982-13  
; Sequence 13, Application US/08766982  
; Patent No. 5948892  
; GENERAL INFORMATION:  
; APPLICANT: Wahl, Robert C.  
; TITLE OF INVENTION: Analogs of Macrophage Stimulating  
; TITLE OF INVENTION: Protein  
; NUMBER OF SEQUENCES: 13

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,982
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-441
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-766-982-13

Query Match 60.6%; Score 702; DB 2; Length 224;
Best Local Similarity 60.3%; Pred. No. 1.4e-72;
Matches 135; Conservative 26; Mismatches 59; Indels 4; Gaps 3;

QY 1 IVGVETXHSQAHQVSLNSGYHXCGLSLNXXWVSAAHCKYKSLVRLGHHIRVNEG 60
Db 1 IVGVYNCSENSVPQVSLNSGYHFCGSLNQWVVSAGHYKSRQVRLGHHIEVLEG 60
QY 61 TEQXISSSVXRHPNYSVNIKNDIMLIKLTYPATLQYVHVALPTECAADATMCTVSG 120
Db 61 NEQFINAKIIRHPQVDRKTLNNDIMLIKLSRAVINARVSTISLPTAPATGKCLISG 120
QY 121 WNTMTSSVXD-GDKLQXILPILSHADCANSGPGMITQSMFCAGYLEGKDSQCGDSGG 179
Db 121 WNTMTSSVXD-GDKLQXILPILSHADCANSGPGMITQSMFCAGYLEGKDSQCGDSGG 179
QY 180 PVVNCNGVLQGVVSWGVCGERDXPGVWAKVXVLSGWRDTMA 223
Db 180 PVVNCNGVLQGVVSWGVCGERDXPGVWAKVXVLSGWRDTMA 223

RESULT 9
US-08-944-483-36
; Sequence 36, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: PASESO for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,483
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6183.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6232456e
; US-08-944-483-36

Query Match 60.6%; Score 702; DB 4; Length 224;
Best Local Similarity 60.3%; Pred. No. 1.4e-72;
Matches 135; Conservative 26; Mismatches 59; Indels 4; Gaps 3;

QY 1 IVGVETXHSQAHQVSLNSGYHXCGLSLNXXWVSAAHCKYKSLVRLGHHIRVNEG 60
Db 1 IVGVYNCSENSVPQVSLNSGYHFCGSLNQWVVSAGHYKSRQVRLGHHIEVLEG 60
QY 61 TEQXISSSVXRHPNYSVNIKNDIMLIKLTYPATLQYVHVALPTECAADATMCTVSG 120
Db 61 NEQFINAKIIRHPQVDRKTLNNDIMLIKLSRAVINARVSTISLPTAPATGKCLISG 120
QY 121 WNTMTSSVXD-GDKLQXILPILSHADCANSGPGMITQSMFCAGYLEGKDSQCGDSGG 179
Db 121 WNTMTSSVXD-GDKLQXILPILSHADCANSGPGMITQSMFCAGYLEGKDSQCGDSGG 179
QY 180 PVVNCNGVLQGVVSWGVCGERDXPGVWAKVXVLSGWRDTMA 223
Db 180 PVVNCNGVLQGVVSWGVCGERDXPGVWAKVXVLSGWRDTMA 223

RESULT 10
US-09-296-219-13
; Sequence 13, Application US/09296219
; Patent No. 6248560
; GENERAL INFORMATION:
; APPLICANT: Wahl, Robert C.
; TITLE OF INVENTION: Analogs of Macrophage Stimulating
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/296,219

```

```

; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-441
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-296-219-13

Query Match 60.6%; Score 702; DB 4; Length 224;
Best Local Similarity 60.3%; Pred. No. 1.4e-72;
Matches 135; Conservative 26; Mismatches 59; Indels 4; Gaps 3;

QY 1 IVGGYECTXHSQAQVSLNSGYHXCGLSLNXXWVYSAACHYKSVLRVRLGHEHHRVNEG 60
DB 1 IVGGYCNCEENSPYQVSLNSGYHFCGSLINEQWVVSAGHCYKSRIOVRLGHEHIEVLEG 60
QY 61 TEQXISSSVXRHPNYSSYNIXNDIMLIKLTYPATLNQYVHAVALPTECAADATMCTVSG 120
DB 61 NEQFINAAKIIRHPQYDRKTLNNDIMLIKLSRAVINARVSTISLPTAPPATGTKCLISG 120
QY 121 WGMTSMSSVXD-GDKLOXLPLILSHADCANSGYGMITQSMFCAGYLEGGKDCSCQDSSG 179
DB 121 WGMTASSGADYPDELQCLDAPVLSQAKCEASY-PGKITSNMFCVGFLEGGKDCSCQDSSG 179
QY 180 PVVNCNGVLQGVVSWGYGCAERDXPGVYAKVXVLSGWVRDTMA 223
DB 180 PVVNCNGQLQ--GVVSWGDCGAQKNKPGVYTKVYVYVVKWIKNTIA 221

RESULT 11
US-09-027-337-5
; Sequence 5, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: Tanimoto, Hirotooshi
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in Breast and Ovarian Carcinomas
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 5
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Serine protease catalytic domain of trypsin (Try)
; US-09-027-337-5

Query Match 60.6%; Score 702; DB 2; Length 225;
Best Local Similarity 60.3%; Pred. No. 1.4e-72;
Matches 135; Conservative 26; Mismatches 59; Indels 4; Gaps 3;

QY 1 IVGGYECTXHSQAQVSLNSGYHXCGLSLNXXWVYSAACHYKSVLRVRLGHEHHRVNEG 60
DB 2 IVGGYCNCEENSPYQVSLNSGYHFCGSLINEQWVVSAGHCYKSRIOVRLGHEHIEVLEG 61
QY 61 TEQXISSSVXRHPNYSSYNIXNDIMLIKLTYPATLNQYVHAVALPTECAADATMCTVSG 120
DB 62 NEQFINAAKIIRHPQYDRKTLNNDIMLIKLSRAVINARVSTISLPTAPPATGTKCLISG 121
QY 121 WGMTSMSSVXD-GDKLOXLPLILSHADCANSGYGMITQSMFCAGYLEGGKDCSCQDSSG 179
DB 122 WGMTASSGADYPDELQCLDAPVLSQAKCEASY-PGKITSNMFCVGFLEGGKDCSCQDSSG 180

; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-441
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-296-219-13

Query Match 60.6%; Score 702; DB 4; Length 224;
Best Local Similarity 60.3%; Pred. No. 1.4e-72;
Matches 135; Conservative 26; Mismatches 59; Indels 4; Gaps 3;

QY 1 IVGGYECTXHSQAQVSLNSGYHXCGLSLNXXWVYSAACHYKSVLRVRLGHEHHRVNEG 60
DB 1 IVGGYCNCEENSPYQVSLNSGYHFCGSLINEQWVVSAGHCYKSRIOVRLGHEHIEVLEG 60
QY 61 TEQXISSSVXRHPNYSSYNIXNDIMLIKLTYPATLNQYVHAVALPTECAADATMCTVSG 120
DB 61 NEQFINAAKIIRHPQYDRKTLNNDIMLIKLSRAVINARVSTISLPTAPPATGTKCLISG 120
QY 121 WGMTSMSSVXD-GDKLOXLPLILSHADCANSGYGMITQSMFCAGYLEGGKDCSCQDSSG 179
DB 121 WGMTASSGADYPDELQCLDAPVLSQAKCEASY-PGKITSNMFCVGFLEGGKDCSCQDSSG 179
QY 180 PVVNCNGVLQGVVSWGYGCAERDXPGVYAKVXVLSGWVRDTMA 223
DB 180 PVVNCNGQLQ--GVVSWGDCGAQKNKPGVYTKVYVYVVKWIKNTIA 221

RESULT 12
US-09-644-600-5
; Sequence 5, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in Carcinomas
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 5
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Trypsin
; US-09-644-600-5

Query Match 60.6%; Score 702; DB 4; Length 225;
Best Local Similarity 60.3%; Pred. No. 1.4e-72;
Matches 135; Conservative 26; Mismatches 59; Indels 4; Gaps 3;

QY 1 IVGGYECTXHSQAQVSLNSGYHXCGLSLNXXWVYSAACHYKSVLRVRLGHEHHRVNEG 60
DB 2 IVGGYCNCEENSPYQVSLNSGYHFCGSLINEQWVVSAGHCYKSRIOVRLGHEHIEVLEG 61
QY 61 TEQXISSSVXRHPNYSSYNIXNDIMLIKLTYPATLNQYVHAVALPTECAADATMCTVSG 120
DB 62 NEQFINAAKIIRHPQYDRKTLNNDIMLIKLSRAVINARVSTISLPTAPPATGTKCLISG 121
QY 121 WGMTSMSSVXD-GDKLOXLPLILSHADCANSGYGMITQSMFCAGYLEGGKDCSCQDSSG 179
DB 122 WGMTASSGADYPDELQCLDAPVLSQAKCEASY-PGKITSNMFCVGFLEGGKDCSCQDSSG 180
QY 180 PVVNCNGVLQGVVSWGYGCAERDXPGVYAKVXVLSGWVRDTMA 223
DB 181 PVVNCNGQLQ--GVVSWGDCGAQKNKPGVYTKVYVYVVKWIKNTIA 222

RESULT 13
US-08-278-091-9
; Sequence 9, Application US/08278091
; Patent No. 5506139
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```



```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-493 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-472-173-9

Query Match 60.5%; Score 701; DB 1; Length 223;
Best Local Similarity 60.3%; Pred. No. 1.8e-72;
Matches 135; Conservative 29; Mismatches 56; Indels 4; Gaps 3;

QY 1 IVGGYECTHSHQAHQVSLNSGYHXCGLSLINXXWVSAAHCYKSVLRVLGEHHIRVNEG 60
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1 IVGGYTCGANTVPYQVSLNSGYHFCGGLSLNSQWVSAAHCYKSGIQVRLGEDNINVEG 60
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 TEQXISSSVYRHPNYSYNIKXNDIMLIKLTXPATLNQYVHAVALPTECAADATMCTVSG 120
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 NEQFISASKSIVHPNSYNSNTLNNDIMLIKLSAASLSRSVASISLPTSCASAGTQCLISG 120
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 WGNMTSS-VXDGDKLQXLPLSHADCANSTQSGMITQSMFCAGYLEGGKDSQGDSSG 179
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 WGNTKSGTSYPDVLKCLKAPILSDSSCKSAY-PQITSNMFCAGYLEGGKDSQGDSSG 179
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 180 PIVCNGLVQGVVSWGYGCAERDXPGVYAKVXVLSGWRDTMA 223
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 180 PIVCGKIQ--GIVSWGCGCAQKNKPGVYTKVNVSWIKQTIA 221
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: February 12, 2003, 10:30:12  
Job time : 49.3582 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:17:50 ; Search time 56.4179 Seconds  
(without alignments)  
383.393 Million cell updates/sec

Title: US-10-036-371-8

Perfect score: 1158

Sequence: 1 IVGGVECTXHSQAHQVLSLNS.....GVYAKVXVLSGWVDRDTMAXY 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 73.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1086.5	93.8	240	2 S39047	trypsin (EC 3.4.21
2	1085.5	93.7	241	2 S39048	trypsin (EC 3.4.21
3	905.5	78.2	242	2 S31776	trypsin (EC 3.4.21
4	901.5	77.8	242	2 S31775	trypsin (EC 3.4.21
5	897.5	77.5	242	2 S49489	trypsin (EC 3.4.21
6	893.5	77.2	231	2 S31778	trypsin (EC 3.4.21
7	758.5	65.5	229	1 TRDPS	trypsin (EC 3.4.21
8	756	65.3	228	2 S31779	trypsin (EC 3.4.21
9	755	65.2	246	1 TRRT1	trypsin (EC 3.4.21
10	729	63.0	246	2 B25528	trypsin (EC 3.4.21
11	727	62.8	247	1 TRDG	trypsin (EC 3.4.21
12	726	62.7	246	1 TRRT2	trypsin (EC 3.4.21
13	721	62.3	247	2 S13813	trypsin (EC 3.4.21
14	719	62.1	247	2 A27547	trypsin (EC 3.4.21
15	718	62.0	243	2 A35871	trypsin (EC 3.4.21
16	710	61.3	247	1 B25852	trypsin (EC 3.4.21
17	707	61.1	231	1 TRPGTR	trypsin (EC 3.4.21
18	706	61.0	246	1 TRDGC	trypsin (EC 3.4.21
19	702	60.6	247	1 A25852	trypsin (EC 3.4.21
20	701	60.5	229	1 TRBOTR	trypsin (EC 3.4.21
21	696	60.1	248	2 S55067	trypsin (EC 3.4.21
22	691	59.7	304	2 S33496	trypsin (EC 3.4.21
23	681.5	58.9	259	2 I38363	trypsin (EC 3.4.21
24	674	58.2	248	2 S55066	trypsin (EC 3.4.21
25	665	57.4	247	2 S12764	trypsin (EC 3.4.21
26	650.5	56.2	247	2 S05494	trypsin (EC 3.4.21
27	617.5	53.3	246	2 J01472	trypsin (EC 3.4.21
28	616.5	53.2	246	2 J01471	trypsin (EC 3.4.21
29	566	48.9	250	2 T01779	trypsin (EC 3.4.21

30	529	45.7	250	2 S31384	trypsin (EC 3.4.21
31	474	40.9	266	2 S54146	trypsin (EC 3.4.21
32	471.5	40.7	275	2 S40007	trypsin (EC 3.4.21
33	467.5	40.4	274	2 S35339	trypsin (EC 3.4.21
34	467.5	40.4	277	2 S35340	trypsin (EC 3.4.21
35	458	39.6	267	2 S40006	trypsin (EC 3.4.21
36	449.5	38.8	263	2 S47537	chymotrypsin (EC 3
37	447.5	38.6	261	1 NGMSG	7S nerve growth fa
38	446.5	38.6	261	2 A29745	tissue kallikrein
39	441.5	38.1	261	2 A29586	tissue kallikrein
40	435	37.6	232	1 KQPG	tissue kallikrein
41	433.5	37.4	244	2 S72219	chymotrypsin B - A
42	431	37.2	237	1 TRCY1	trypsin (EC 3.4.21
43	431	37.2	253	2 A53968	serine proteinase
44	428.5	37.0	254	1 TRW3Y	trypsin-like prote
45	427.5	36.9	244	2 A44284	tissue kallikrein

ALIGNMENTS

RESULT 1

S39047

trypsin (EC 3.4.21.4) I - Atlantic cod

C:Species: Gadus morhua (Atlantic cod)

C:Date: 20-May-1994 #sequence\_revision 10-Nov-1995 #text\_change 20-Jun-2000

C:Accession: S39047

R:Gudmundsdottir, A.; Gudmundsdottir, E.; Oskarsson, S.; Bjarnason, J.B.; Eakin, A.K.; C

Eur. J. Biochem. 217, 1091-1097, 1993

A:Title: Isolation and characterization of cDNAs from Atlantic cod encoding two different

A:Reference number: S39047; MUID:94039130; PMID:8223632

A:Accession: S39047

A:Molecule type: mRNA

A:Residues: 1-240 <GUD>

A:Cross-references: EMBL:X76886; NID:g450517; PIDN:CAA54214.1; PID:gl334752

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; serine proteinase

F:19-233/Domain: trypsin homology <TRY>

F:58,102,194/Active site: His, Asp, Ser #status predicted

Query Match 93.8%; Score 1086.5; DB 2; Length 240;  
Best Local Similarity 91.6%; Pred. No. 4.9e-99;  
Matches 206; Conservative 2; Mismatches 14; Indels 3; Gaps 2;

Qy 1 IVGGVECTXHSQAHQVLSLNSGVHFCGGSLVSKDWVVSAAHCYKSVLRVLGEHHRVNEG 60

Db 19 IVGGVECTXHSQAHQVLSLNSGVHFCGGSLVSKDWVVSAAHCYKSVLRVLGEHHRVNEG 78

Qy 61 TEQXISSSVYKRPNPYSSVYINXNDIMLIKLTXPATLNQYHVALPTECAADATWCTVSG 120

Db 79 TEQYISSSVIRHPNPYSSVYINXNDIMLIKLTXPATLNQYHVALPTECAADATWCTVSG 138

Qy 121 WGNMTSSVYXGDKLQKXLPILSHADCANSYFGQMITQSMFCAGYLEGGKDCSQDSSGP 180

Db 139 WGNMTSSVYADGDKLQKXLPILSHADCANSY-PGMITQSMFCAGYLEGGKDCSQDSSGP 197

Qy 181 VVCGNLQGVGVSWGVSGYCAERDPGVYAKVXVLSGWVDRDTMAXY 225

Db 198 VVCGNLQ- GVVSWSYCGCAERDPGVYAKVXVLSGWVDRDTMAXY 240

RESULT 2

S39048

trypsin (EC 3.4.21.4) X - Atlantic cod

C:Species: Gadus morhua (Atlantic cod)

C:Date: 20-May-1994 #sequence\_revision 03-Aug-1995 #text\_change 20-Jun-2000

C:Accession: S39048

R:Gudmundsdottir, A.; Gudmundsdottir, E.; Oskarsson, S.; Bjarnason, J.B.; Eakin, A.K.; C

Eur. J. Biochem. 217, 1091-1097, 1993

A:Title: Isolation and characterization of cDNAs from Atlantic cod encoding two different

A:Reference number: S39047; MUID:94039130; PMID:8223632

A:Accession: S39048

A>Status: preliminary





```

Best Local Similarity 75.1%; Pred. No. 1.7e-80;
Matches 169; Conservative 16; Mismatches 37; Indels 3; Gaps 2;

QY 1 IVGGYECTXHSQAHQVSLNSGVHXCGLSLINXWVWSAAHCYKSVLRVRLGHEHHIRVNEG 60
DB 21 IVGGYECTXHSQAHQVSLNSGVHXCGLSLINXWVWSAAHCYKSVLRVRLGHEHHIRVNEG 80

QY 61 TEQXISSSVXRHPNYSYNYNDIMLKLTXPATLNQYVHAVALPTECAADATMCTVSG 120
DB 81 KEQFISSSRVIRHPNYSYNYNDIMLKLTXPATLNQYVHAVALPTECAADATMCTVSG 140

QY 121 WGNMTSSVXDGDKLQXLPILSHADCANSGPGMITOSMFCAGYLEGGKDCSCQDSGGP 180
DB 141 WGNMTSSVXDGDKLQXLPILSHADCANSGPGMITOSMFCAGYLEGGKDCSCQDSGGP 199

QY 181 VVCGVQLQGVVWSWGYGCAERDPGVYAKVXVLSGWVDRDTMAXY 225
DB 200 VVCGVQLQ--GVWSWGYGCAERDPGVYAKVXVLSGWVDRDTMAXY 242

RESULT 6
S31778
trypsin (EC 3.4.21.4) II precursor - Atlantic salmon (fragment)
C:Species: Salmo salar (Atlantic salmon)
C>Date: 22-Nov-1993 #sequence_revision 03-Aug-1995 #text_change 22-Jun-1999
C:Accession: S66657; S31778
R:Male, R.; Lorenz, J.B.; Smalas, A.O.; Torrisen, K.R.
A:Title: Molecular cloning and characterization of anionic and cationic variants of trypsin
A:Reference number: S66657; PMID:96035908; PMID:7556223
A:Accession: S66658
A:Molecule type: mRNA
A:Residues: 1-231 <MAL>
A:Cross-references: EMBL:X70073; NID:G64385; PIDN:CAA49678.1; PID:G64386
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:1-4/Domain: signal sequence (fragment) #status predicted <SIG>
F:5-9/Domain: activation peptide #status predicted <APT>
F:10-231/Product: trypsin II #status predicted <MAT>
F:10-224/Domain: trypsin homology <TRY>
F:16-145,34-50,118-218,125-191,156-170,181-205/Disulfide bonds: #status predicted
F:49,93,185/Active site: His, Asp, Ser #status predicted

Query Match 77.2%; Score 893.5; DB 2; Length 231;
Best Local Similarity 74.7%; Pred. No. 3.9e-80;
Matches 168; Conservative 16; Mismatches 38; Indels 3; Gaps 2;

QY 1 IVGGYECTXHSQAHQVSLNSGVHXCGLSLINXWVWSAAHCYKSVLRVRLGHEHHIRVNEG 60
DB 10 IVGGYECTXHSQAHQVSLNSGVHXCGLSLINXWVWSAAHCYKSVLRVRLGHEHHIRVNEG 69

QY 61 TEQXISSSVXRHPNYSYNYNDIMLKLTXPATLNQYVHAVALPTECAADATMCTVSG 120
DB 70 SEQFISSSRVIRHPNYSYNYNDIMLKLTXPATLNQYVHAVALPTECAADATMCTVSG 129

QY 121 WGNMTSSVXDGDKLQXLPILSHADCANSGPGMITOSMFCAGYLEGGKDCSCQDSGGP 180
DB 130 WGNMTSSVXDGDKLQXLPILSHADCANSGPGMITOSMFCAGYLEGGKDCSCQDSGGP 188

QY 181 VVCGVQLQGVVWSWGYGCAERDPGVYAKVXVLSGWVDRDTMAXY 225
DB 189 VVCGVQLQ--GVWSWGYGCAERDPGVYAKVXVLSGWVDRDTMAXY 231

RESULT 7
TRDS
trypsin (EC 3.4.21.4) precursor - spiny dogfish
N:Alternate names: trypsinogen
C:Species: Squalus acanthias (spiny dogfish)
C>Date: 31-May-1979 #sequence_revision 31-May-1979 #text_change 21-Jan-1997
C:Accession: A00950; B27719
R:Titani, K.; Ericsson, L.H.; Neurath, H.; Walsh, K.A.
Biochemistry 14, 1358-1366, 1975

```

```

A:Title: Amino acid sequence of dogfish trypsin.
A:Reference number: A00950; PMID:75146445; PMID:1092332
A:Accession: A00950
A:Molecule type: protein
A:Residues: 8-229 <TIT>
A>Note: 119-Pro was also found
R:Hermanson, M.A.; Tye, R.W.; Reese, G.R.; Neurath, H.; Walsh, K.A.
FBS Lett. 14, 222-224, 1971
A:Title: Comparison of the amino terminal sequences of bovine, dogfish, and lungfish try
A:Reference number: A27719
A:Accession: B27719
A:Molecule type: protein
A:Residues: 1-21 <HER>
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; polymorphism; protein digestion; serine proteinase; zym
F:1-7/Domain: activation peptide #status experimental <APT>
F:8-229/Product: trypsin #status predicted <MAT>
F:8-222/Domain: trypsin homology <TRY>
F:14-143,32-48,116-216,123-189,154-168,179-203/Disulfide bonds: #status predicted
F:47,91,183/Active site: His, Asp, Ser #status predicted
F:59,61,64,69/Binding site: calcium (Glu, Asp, Ala, Glu) #status predicted

Query Match 65.5%; Score 758.5; DB 1; Length 229;
Best Local Similarity 62.8%; Pred. No. 6.7e-67;
Matches 140; Conservative 23; Mismatches 57; Indels 3; Gaps 2;

QY 1 IVGGYECTXHSQAHQVSLNSGVHXCGLSLINXWVWSAAHCYKSVLRVRLGHEHHIRVNEG 60
DB 8 IVGGYECTXHSQAHQVSLNSGVHXCGLSLINXWVWSAAHCYKSVLRVRLGHEHHIRVNEG 67

QY 61 TEQXISSSVXRHPNYSYNYNDIMLKLTXPATLNQYVHAVALPTECAADATMCTVSG 120
DB 68 DEYIYSSSVIRHPNYSYNYNDIMLKLTXPATLNQYVHAVALPTECAADATMCTVSG 127

QY 121 WGNMTSSVXDGDKLQXLPILSHADCANSGPGMITOSMFCAGYLEGGKDCSCQDSGGP 180
DB 128 WGNMTDGAIVSGDQLCLDAPVLSDAECKGAY-PCGITNNMVCVWYEGKDCSCQDSGGP 186

QY 181 VVCGVQLQGVVWSWGYGCAERDPGVYAKVXVLSGWVDRDTMAXY 223
DB 187 VVCGVQLQ--GVWSWGYGCAERDPGVYAKVXVLSGWVDRDTMAXY 227

RESULT 8
S31779
trypsin (EC 3.4.21.4) III precursor - Atlantic salmon (fragment)
C:Species: Salmo salar (Atlantic salmon)
C>Date: 22-Nov-1993 #sequence_revision 03-Aug-1995 #text_change 22-Jun-1999
C:Accession: S66657; S31779
R:Male, R.; Lorenz, J.B.; Smalas, A.O.; Torrisen, K.R.
Eur. J. Biochem. 232, 677-685, 1995
A:Title: Molecular cloning and characterization of anionic and cationic variants of tryp
A:Reference number: S66657; PMID:96035908; PMID:7556223
A:Accession: S66657
A:Molecule type: mRNA
A:Residues: 1-238 <MAL>
A:Cross-references: EMBL:X70074; NID:G64387; PIDN:CAA49679.1; PID:G64388
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:1-7/Domain: signal sequence (fragment) #status predicted <SIG>
F:8-15/Domain: activation peptide #status predicted <APT>
F:16-238/Product: trypsin III #status predicted <MAT>
F:16-231/Domain: trypsin homology <TRY>
F:22-152,40-56,124-225,131-198,163-177,188-212/Disulfide bonds: #status predicted
F:55,99,192/Active site: His, Asp, Ser #status predicted

Query Match 65.3%; Score 756; DB 2; Length 238;
Best Local Similarity 63.8%; Pred. No. 1.2e-66;
Matches 143; Conservative 26; Mismatches 51; Indels 4; Gaps 3;

QY 1 IVGGYECTXHSQAHQVSLNSGVHXCGLSLINXWVWSAAHCYKSVLRVRLGHEHHIRVNEG 60
DB 16 IVGGYECTXHSQAHQVSLNSGVHXCGLSLINXWVWSAAHCYKSVLRVRLGHEHHIRVNEG 75

```

```

Qy   61 TEQXISSSVKRPNTSSYNIKNIDIMLIKLTXPATLNQVYHVAVALPTECAADTMCTVSG 120
Db   76 TEQFIDSVKVWMPSPNSNRNLDNDIMLIKLSKPASLNSYSTVALPSSCASSCTRCCLVSG 135
Qy   121 WGNMTSMSSVD-GRKLOXLQLPILSHADCANSYGPGMITOSMFCAGYLEGGKDSCQDSSG 179
Db   136 WGNLSGSSNPDTLRCLDLPLSSSSNCNAY-PQGITSNNFCAGFMWEGGDKSCQDSSG 194
Qy   180 PVVCNGVLQGVVGWSWGCGAERDXPGVYAKVXVLSGWVRDTWA 223
Db   195 PVVCNGQLQ--GVVSWGYGCAQRNKPGVYTKVCYNRSWISWTMS 236

RESULT 9
TRRT1
N:trypsin (EC 3.4.21.4) I precursor - rat
N:Alternate names: trypsinogen I
C:Species: Rattus norvegicus (Norway rat)
C>Date: 17-Dec-1982 #sequence_revision 17-Dec-1982 #text_change 24-Sep-1989
C:Accession: B22657; A00948
R:Craik, C.S.; Choo, Q.L.; Swift, G.H.; Quinto, C.; MacDonald, R.J.; Rutter, W.J.
J: Biol. Chem. 259, 14255-14264, 1984
A:Title: Structure of two related rat pancreatic trypsin genes.
A:Reference number: A22657; MUID:85054880; PMID:6094547
A:Accession: B22657
A:Molecule type: DNA
A:Residues: 1-246 <CR>
A:Cross-references: GB:J00778; NID:g206507; PIDN:AAA98518.1; PID:g206508
A>Note: The authors translated the codon ATC for residue 6 as Leu and GAC for residue 17 as Arg.
R:MacDonald, R.J.; Stary, S.J.; Swift, G.H.
J: Biol. Chem. 257, 9724-9732, 1982
A:Title: Two similar but nonallelic rat pancreatic trypsins. Nucleotide sequences of cDNA clones.
A:Reference number: A00948; MUID:82265624; PMID:6896710
A:Accession: A00948
A:Molecule type: mRNA
A:Residues: 1-246 <MAC>
A:Cross-references: GB:J00778; NID:g206507; PIDN:AAA98518.1; PID:g206508
C:Genetics:
A:Intron: 14/1; 67/2; 152/1; 197/3
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-23/Domain: activation peptide #status predicted <APT>
F:24-246/Product: trypsin I #status predicted <ENZ>
F:24-239/Domain: trypsin homology <TRY>
F:30-160, 48-64, 132-233, 139-206, 171-185/Dialfide bonds: #status predicted
F:63, 107, 200/Active site: His, Asp, Ser #status predicted
F:75, 77, 80, 85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match      65.2%; Score 755; DB 1; Length 246;
Best Local Similarity 63.8%; Pred. No. 1.6e-66;
Matches 143; Conservative 25; Mismatches 53; Indels 4; Gaps 3;

Qy   1 IVGYEECTXHSQAHQVSLNSGTHXCGLSINXXWTVVSAHCYSVLRGLCEHHIRVNNEG 60
Db   24 IVGYTCPEHSVPYQVSLNSGHYFCGSLINDQWVVSAHCYSRIQRLRGHNINVLEG 83
Qy   61 TEQXISSSVKRPNTSSYNIKNIDIMLIKLTXPATLNQVYHVAVALPTECAADTMCTVSG 120
Db   84 DEQFINAAKIIRPNSTTSWLNDNDIMLIKLSFPVKLNARVAPVALPSACAPAGTQCCLISG 143
Qy   121 WGNMTSMSSVD-GRKLOXLQLPILSHADCANSYGPGMITOSMFCAGYLEGGKDSCQDSSG 179
Db   144 WGNLTSLNGVNNPDLQCVDAPVLSQADCEAAY-PGEITSSMICVGFLEGGKDSCQDSSG 202
Qy   180 PVVCNGVLQGVVGWSWGCGAERDXPGVYAKVXVLSGWVRDTWA 223
Db   203 PVVCNGQLQ--GVVSWGYGCALPDNFGVYTKVCNFWGIQDTIA 244

RESULT 10
B25528

```

```

trypsin (EC 3.4.21.4) precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 22-Jun-1999
C:Accession: B25528
R:Stevenson, B.J.; Hagenbuechle, O.; Wellauer, P.K.
Nucleic Acids Res. 14, 8307-8330, 1986
A:Title: Sequence organisation and transcriptional regulation of the mouse elastase II ar
A:Reference number: A93646; MUID:87066713; PMID:3641189
A:Reference number: A93646; MUID:87066713; PMID:3641189
A:Accession: B25528
A:Molecule type: mRNA
A:Residues: 1-246 <STE>
A:Cross-references: GB:X04574; NID:g54918; PIDN:CAA28243.1; PID:g54919
C:Superfamily: trypsin; trypsin homology
C:Keywords: calcium binding; hydrolase; protein digestion; serine proteinase
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-246/Product: trypsin #status predicted <WAT>
F:24-239/Domain: trypsin homology <TRY>
F:30-160, 48-64, 132-233, 139-206, 171-185/Disulfide bonds: #status predicted
F:63, 107, 200/Active site: His, Asp, Ser #status predicted
F:75, 77, 80, 85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 63.0%; Score 729; DB 2; Length 246;
Best Local Similarity 61.2%; Pred. No. 5,7e-64;
Matches 137; Conservative 27; Mismatches 56; Indels 4; Gaps 3;

Qy 1 IVGGYECTYHSQAHQVSLNSGYHXCGLSLNXXWVYSAACHYKSVLRVRLGEHHIRVNEG 60
Db 24 IVGGYTCRESSVPYQVSLNAGYHFCGSLINDQWVYSAACHYKSVLRVRLGEHHIRVNEG 83

Qy 61 TEQXISSSSVXRHPNYSYINXNDIMLIKLTXPATLNQYHVALPTECAADATMCTVSG 120
Db 84 NEQFVDSAKIIRHPNYSNWTLDNIMLIKASPVTLNARVASVPLPSSCAPAGTQCLISG 143

Qy 121 WGNWTSS-VYDQKQLXLLPILSHADCANSYGPMTQTSMFCAGYLEGGKDSQCGDSGG 179
Db 144 WGNLTSLNGVNNPDLQCVDPAPVLPQADCEASY-PGDTINNMICVGFLEGGKDSQCGDSGG 202

Qy 180 PIVCNGLQGVGVSWGYGCAERDFGVYAKVXVLSCGVDRDTMA 223
Db 203 PIVCNGLQ--GIVSWGYGCAQDPAGVYTKVCNVYVDWIGNTIA 244

RESULT 11
TRDG
trypsin (EC 3.4.21.4) precursor, anionic - dog
N:Alternate names: cationic trypsinogen
C:Species: Canis lupus familiaris (dog)
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 18-Jun-1999
C:Accession: A26273
R:Pinsky, S.D.; LaForge, K.S.; Scheele, G.
Mol. Cell. Biol. 5, 2669-2676, 1985
A:Title: Differential regulation of trypsinogen mRNA translation: full-length mRNA sequence
A:Reference number: A26273; MUID:86284628; PMID:3841794
A:Accession: A26273
A:Molecule type: mRNA
A:Residues: 1-247 <PIN>
A:Cross-references: GB:M11589; NID:g164094; PIDN:AAA30899.1; PID:g164095
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-23/Domain: activation peptide #status predicted <APT>
F:24-247/Product: trypsin, anionic #status predicted <ENZ>
F:24-239/Domain: trypsin homology <TRY>
F:30-160, 48-64, 132-233, 139-206, 171-185/Disulfide bonds: #status predicted
F:63, 107, 200/Active site: His, Asp, Ser #status predicted
F:75, 77, 80, 85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 62.8%; Score 727; DB 1; Length 247;
Best Local Similarity 61.6%; Pred. No. 9e-64;
Matches 138; Conservative 31; Mismatches 51; Indels 4; Gaps 3;

Qy 1 IVGGYECTYHSQAHQVSLNSGYHXCGLSLNXXWVYSAACHYKSVLRVRLGEHHIRVNEG 60

```

```
Db 24 IVGGYTCENSVYPQVSLNAGYHFCGSLSDQWVSAAHCKYSRIQVRLGEYNIDVLEGG 83
QY 61 TEQXISSSVXRHPNYSYNIKNDIMLKLTYPATLNQYVHAVALPTCAADATMCTVSG 120
Db 84 NEQFINSAKVIHRPNYSWLDNDIMLKLSPAVLNARVATISLPACAPGTQCLISG 143
QY 121 WNTMSSVVD-GDKLQXLPILSHADCANSGPGMITQSMFCAGYLEGGKDCSQDSSG 179
Db 144 WNTLSSGPNYPQLDAPILTQAQCEASY-PGQITENMIKAGFLGGKDCSQDSSG 202
QY 180 PWCNGVGLQGVGWSGVCABRDXPVYAKVXVLSGWVRDTMA 223
Db 203 PWCNGELQ--GIVSWGCGCAQKPGYTKVCNVYDWTQSTIA 244
RESULT 12
TRRT2
trypsin (EC 3.4.21.4) II precursor - rat
N:Alternate names: trypsinogen II
C:Species: Rattus norvegicus (Norway rat)
C>Date: 05-Apr-1983 #sequence_revision 30-Sep-1987 #text_change 18-Jul-1997
C:Accession: A22657; A00949
R:Craik, C.S.; Choo, Q.L.; Swift, G.H.; Quinto, C.; MacDonald, R.J.; Rutter, W.J.
J. Biol. Chem. 259, 14255-14264, 1984
A:Title: Structure of two related rat pancreatic trypsin genes.
A:Reference number: A22657; MUID:85054880; PMID:6094547
A:Accession: A22657
A:Molecule type: DNA
A:Residues: 1-246 <CR>
R:MacDonald, R.J.; Swift, S.J.; Swift, G.H.
J. Biol. Chem. 257, 9724-9732, 1982
A:Title: Two similar but nonallelic rat pancreatic trypsinogens. Nucleotide sequences of
A:Reference number: A00948; MUID:82265624; PMID:6896710
A:Accession: A00949
A:Molecule type: mRNA
A:Residues: 9-246 <MAC>
C:Comment: The trypsin II mRNA is present in much lower quantities than the trypsin I mRNA.
C:Genetics:
A:Introns: 14/1; 67/2
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-23/Domain: activation peptide #status predicted <APT>
F:24-246/Product: trypsin II #status predicted <BNZ>
F:24-239/Domain: trypsin homology <TRY>
F:30-160,48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted
F:63,107,200/Active site: His, Asp, Ser #status predicted
F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted
Query Match 62.7%; Score 726; DB 1; Length 246;
Best Local Similarity 62.1%; Pred. No. 1.1e-63;
Matches 139; Conservative 24; Mismatches 57; Indels 4; Gaps 3;
QY 1 IVGGYECTXHSQAQHVSLNSGYHXCGLSLINXWVSAAHCKYSVLRVLGEHHRVNEG 60
Db 24 IVGGYTCENSVYPQVSLNAGYHFCGSLSDQWVSAAHCKYSRIQVRLGEHHRVNEG 83
QY 61 TEQXISSSVXRHPNYSYNIKNDIMLKLTYPATLNQYVHAVALPTCAADATMCTVSG 120
Db 84 DEQFINAAKIIKHPNFRKTLNNDIMLKLSPPVKNARVATVALPSSCAPGTQCLISG 143
QY 121 WNTMSS-VXDGDRLQXLPILSHADCANSGPGMITQSMFCAGYLEGGKDCSQDSSG 179
Db 144 WNTLSSGWNPEPDLQCLDAPILLPQADCEASY-PGKIDNMVYCVFLGGKDCSQDSSG 202
QY 180 PWCNGVGLQGVGWSGVCABRDXPVYAKVXVLSGWVRDTMA 223
Db 203 PWCNGELQ--GIVSWGCGCALPDPGYYTKVCNVYDWTQSTIA 244
RESULT 13
SI3813
trypsin (EC 3.4.21.4) - bovine
```

```
C:Species: Bos primigenius taurus (cattle)
C>Date: 02-Dec-1993 #sequence_revision 03-Aug-1995 #text_change 22-Jun-1999
C:Accession: SI3813
R:Le Huron, I.; Wicker, C.; Guilloreau, P.; Toulllec, R.; Puigserver, A.
Eur. J. Biochem. 193, 767-773, 1990
A:Title: Isolation and nucleotide sequence of cDNA clone for bovine pancreatic anionic trypsin.
A:Reference number: SI3813; MUID:91065383; PMID:1701147
A:Accession: SI3813
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-247 <HUB>
A:Cross-references: EMBL:X54703; NID:9829; PIDN:CAA38513.1; PID:9930
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; protein digestion; serine proteinase
F:24-239/Domain: trypsin homology <TRY>
F:63,107,200/Active site: His, Asp, Ser #status predicted
Query Match 62.1%; Score 719; DB 2; Length 247;
Best Local Similarity 61.2%; Pred. No. 5.5e-63;
Matches 137; Conservative 26; Mismatches 57; Indels 4; Gaps 3;
QY 1 IVGGYECTXHSQAQHVSLNSGYHXCGLSLINXWVSAAHCKYSVLRVLGEHHRVNEG 60
Db 25 IVGGYTCOKNSLPYQVSLNAGYHFCGSLINSQWVSAAHCKYSRIQVRLGEHHRVNEG 84
QY 61 TEQXISSSVXRHPNYSYNIKNDIMLKLTYPATLNQYVHAVALPTCAADATMCTVSG 120
Db 85 GEQFIDAAKIIHRPSYNAFTDNDIMLKLSNPATLNSRVTSVLPSCSGSGTKCLVSG 144
QY 121 WNTMSSVVD-GDKLQXLPILSHADCANSGPGMITQSMFCAGYLEGGKDCSQDSSG 179
C:Species: Rattus norvegicus (Norway rat)
C>Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 22-Jun-1999
C:Accession: A27547
R:Fletcher, T.S.; Alhadeff, M.; Craik, C.S.; Largman, C.
Biochemistry 26, 3081-3086, 1987
A:Title: Isolation and characterization of a cDNA encoding rat cationic trypsinogen.
A:Reference number: A27547; MUID:87271609; PMID:3607011
A:Accession: A27547
A:Molecule type: mRNA
A:Residues: 1-247 <FLB>
A:Cross-references: GB:M16624; NID:g206498; PIDN:AAAA1985.1; PID:g206499
C:Superfamily: trypsin; trypsin homology
C:Keywords: calcium binding; hydrolase; protein digestion; serine proteinase
F:25-240/Domain: trypsin homology <TRY>
F:31-161,49-65,133-234,140-207,172-186/Disulfide bonds: #status predicted
F:64,108,201/Active site: His, Asp, Ser #status predicted
F:76,78,81,86/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted
Query Match 62.1%; Score 719; DB 2; Length 247;
Best Local Similarity 61.2%; Pred. No. 5.5e-63;
Matches 137; Conservative 26; Mismatches 57; Indels 4; Gaps 3;
QY 1 IVGGYECTXHSQAQHVSLNSGYHXCGLSLINXWVSAAHCKYSVLRVLGEHHRVNEG 60
Db 25 IVGGYTCOKNSLPYQVSLNAGYHFCGSLINSQWVSAAHCKYSRIQVRLGEHHRVNEG 84
QY 61 TEQXISSSVXRHPNYSYNIKNDIMLKLTYPATLNQYVHAVALPTCAADATMCTVSG 120
Db 85 GEQFIDAAKIIHRPSYNAFTDNDIMLKLSNPATLNSRVTSVLPSCSGSGTKCLVSG 144
QY 121 WNTMSSVVD-GDKLQXLPILSHADCANSGPGMITQSMFCAGYLEGGKDCSQDSSG 179
```

Db 145 WGNLTSLSGTNPSPLLQCLDAPVLSDDSKSSY-PGKITSNMFCLGFLGGKDSQQGDSGG 203  
QY 180 PVVCGVLTQGVVSWGYGCAERDHPGVYAKVXVLSGWRDTMA 223  
Db 204 PVVCGQLQ-GVWSMGYGCAQKRPGVYTKVCNYYNWIIQQTVA 245

RESULT 15  
A35871  
trypsin (EC 3.4.21.4) precursor, pancreatic - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 09-Nov-1990 #sequence\_revision 09-Nov-1990 #text\_change 04-Mar-2000  
C:Accession: A35871; S12117  
R:Shi, Y.B.; Brown, D.D.  
Genes Dev. 4, 1107-1113, 1990  
A:Title: Developmental and thyroid hormone-dependent regulation of pancreatic genes in X  
A:Reference number: A35871; MUID:91007255; PMID:2210372  
A:Accession: A35871  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-243 <SHI>  
A:Cross-references: EMBL:X53458; NID:965162; PIDN:CAA37538.1; PID:965163  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; protein digestion; serine proteinase  
F:1-15/Domain: signal sequence #status predicted <SIG>  
F:16-20/Domain: activation peptide #status predicted <APT>  
F:21-236/Domain: trypsin homology <TRY>  
F:26-243/Product: trypsin I #status predicted <MAT>  
F:27-157,45-61,129-230,136-203,168-182/Disulfide bonds: #status predicted  
F:60,104,197/Active site: His, Asp, Ser #status predicted

Query Match 62.0%; Score 718; DB 2; Length 243;  
Best Local Similarity 61.2%; Pred. No. 6.7e-63;  
Matches 137; Conservative 31; Mismatches 52; Indels 4; Gaps 3;

QY 1 IVGGYECTKHSQAHQVSLNSGYHXCGLSLNXWVSAHCYKSVLRVRLGEHHIRVNEG 60  
Db 21 IIGGATCAKSSVPYIVSLNSGYHFCGSLITNQWVSAHCYKASIQVRLGEHNIALSEG 80

QY 61 TEQXISSSSVXRPHPNVSSYNIXNDIMLIKLTXPATLNOYVHAVALPTECAADATMCTVSG 120  
Db 81 TEQFISSSKVIHSGVNSYTLDDIMLIKLSPASLNAAVNTVPLPSGCSAAGTSCLSIG 140

QY 121 WGNMTSSVXD-GDKLOXLPLILSHADCANSYGPMTQSMFCAGYLEGGKDSQQGDSGG 179  
Db 141 WGNLTSLSGSNYPDLLQCLNAPILITNACNSAY-PGEITANWICVGYNEGKDSQQGDSGG 199

QY 180 PVVCGVLTQGVVSWGYGCAERDHPGVYAKVXVLSGWRDTMA 223  
Db 200 PVVCGQLQ-GVWSMGYGCAMRNYPGVYTKVCNYYNWIIQNTIA 241

Search completed: February 12, 2003, 10:28:56  
Job time : 57.4179 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:04:45 ; Search time 30.2239 Seconds  
(without alignments)  
308.768 Million cell updates/sec

Title: US-10-036-371-8

Perfect score: 1158

Sequence: 1 IVGGVECTXHSQAHQVSLNS.....GVYAKVXVLSGWVDTNAXY 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DR	ID	Description
1	1086.5	93.8	241	1	TRY1_GADMO	P16049 gadus morhu
2	1085.5	93.7	241	1	TRYX_GADMO	Q91041 gadus morhu
3	901.5	77.8	242	1	TRY1_SALSA	P35031 salmo salar
4	893.5	77.2	231	1	TRY2_SALSA	P35032 salmo salar
5	758.5	65.5	229	1	TRY3_SQUAC	P00764 squallus aca
6	756	65.3	238	1	TRY3_SALSA	P35033 salmo salar
7	755	65.2	246	1	TRY1_RAT	P00762 rattus norv
8	735	63.5	244	1	TRY2_XENLA	P07039 xenopus lae
9	729	63.0	246	1	TRY2_MOUSE	P07146 mus musculu
10	727	62.8	247	1	TRY2_CANPA	P06872 canis famil
11	726	62.7	246	1	TRY2_RAT	P00763 rattus norv
12	721	62.3	247	1	TRY2_BOVIN	Q29463 bos taurus
13	719	62.1	247	1	TRY3_RAT	P08426 rattus norv
14	718	62.0	243	1	TRY1_XENLA	P19799 xenopus lae
15	710	61.3	247	1	TRY2_HUMAN	P07478 homo sapien
16	707	61.1	231	1	TRY2_PIG	P00761 sus scrofa
17	706	61.0	246	1	TRY1_CANFA	P06871 canis famil
18	702	60.6	247	1	TRY1_HUMAN	P07477 homo sapien
19	701	60.5	243	1	TRY1_BOVIN	P00760 bos taurus
20	697	60.2	248	1	TRY1_CHICK	Q90627 gallus gall
21	693	59.8	248	1	TRY2_CHICK	Q90628 gallus gall
22	691	59.7	304	1	TRY4_HUMAN	P35030 homo sapien
23	674	58.2	248	1	TRY3_CHICK	Q90629 gallus gall
24	665	57.4	247	1	TRY3_HUMAN	P15951 homo sapien
25	650.5	56.2	246	1	TRY4_RAT	P12788 rattus norv
26	617.5	53.3	246	1	TRYB_RAT	P32822 rattus norv
27	616.5	53.2	246	1	TRYA_RAT	P32821 rattus norv
28	529	45.7	250	1	TRYP_PLEPL	P35034 pleuronecte
29	494	42.7	256	1	KLKB_HUMAN	Q942r5 homo sapien
30	471.5	40.7	275	1	TRY3_ANOGA	P35037 anopheles g
31	467.5	40.4	277	1	TRY2_ANOGA	P35036 anopheles g
32	460.5	39.8	274	1	TRY1_ANOGA	P35035 anopheles g
33	460	39.7	251	1	KLKE_HUMAN	Q9p0g3 homo sapien

34	458	39.6	267	1	TRY7_ANOGA	P35041 anopheles g
35	451	38.9	250	1	KLKB_HUMAN	Q942r5 homo sapien
36	449.5	38.8	263	1	CTRA_GADMO	P47796 gadus morhu
37	448	38.7	244	1	KLK6_HUMAN	Q92876 mus musculu
38	447.5	38.6	261	1	KLK3_MOUSE	P00756 mus musculu
39	446.5	38.6	261	1	KLK9_MOUSE	P15949 mus musculu
40	441.5	38.1	261	1	KLK2_HUMAN	P20151 homo sapien
41	433.5	37.4	245	1	CTRB_GADMO	P06446 gadus morhu
42	433.5	37.4	246	1	KLK_PIG	P00752 sus scrofa
43	432.5	37.3	277	1	KLKD_HUMAN	Q942r3 homo sapien
44	431	37.2	237	1	TRYP_ASTFL	P00765 astacus flu
45	431	37.2	253	1	KLK7_HUMAN	P49862 homo sapien

## ALIGNMENTS

```

RESULT 1
TRY1_GADMO
ID   TRY1_GADMO          STANDARD;          PRT;   241 AA.
AC   P16049; Q91040; Q92156;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DE   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Trypsin I precursor (EC 3.4.21.4).
OS   Gadus morhua (Atlantic cod).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.
OX   NCBI_TaxID=8049;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pyloric caeca.
RX   MEDLINE=94019130; PubMed=8223632;
RA   Gudmundsdottir A., Gudmundsdottir E., Oskarsson S., Bjarnason J.B.,
RA   Rakin A.E., Craik C.S.;
RT   "Isolation and characterization of cDNAs from Atlantic cod encoding
RT   two different forms of trypsinogen.";
RL   Eur. J. Biochem. 217:1091-1097(1993).
RN   [2]
RP   SEQUENCE OF 20-58.
RC   TISSUE=Pyloric caeca.
RX   MEDLINE=89210867; PubMed=2707266;
RA   Asgeirsson B., Fox J.W., Bjarnason J.B.;
RT   "Purification and characterization of trypsin from the poikilotherm
RT   Gadus morhua.";
RL   Eur. J. Biochem. 180:85-94(1989).
CC   -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- SIMILARITY: BELONGS TO TRYPSIN FAMILY SL.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)
-----
EMBL; X76886; CAA54214.1; -.
PIR; S03570; S03570.
HSSP; P00763; IDPO.
MEROPS; S01.151; -.
InterPro; IPR001314; Chymotrypsin.
InterPro; IPR001254; Ser.protease_Try.
Pfam; PF00089; trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
SMART; SM00020; Tryp_SPC; 1.
PROSITE; PS0240; TRYPSIN_DOM; 1.
PROSITE; PS00134; TRYPSIN_HIS; 1.
PROSITE; PS00135; TRYPSIN_SSR; 1.
KW   Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
Multigene family.

```

```

FT SIGNAL 1 13 POTENTIAL.
FT PROPEP 14 19 ACTIVATION PEPTIDE.
FT CHAIN 20 241 TRYPsin I.
FT ACT SITE 59 59 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 103 103 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 195 195 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 26 155 BY SIMILARITY.
FT DISULFID 44 60 BY SIMILARITY.
FT DISULFID 128 228 BY SIMILARITY.
FT DISULFID 135 201 BY SIMILARITY.
FT DISULFID 166 180 BY SIMILARITY.
FT DISULFID 191 215 BY SIMILARITY.
FT SITE 189 189 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
FT CONFLICT 25 25 E -> Q (IN REF. 2).
FT CONFLICT 27 28 TK -> EA (IN REF. 2).
FT CONFLICT 43 43 F -> Y (IN REF. 2).
FT CONFLICT 49 52 VKSD -> IN (IN REF. 2).
SQ SEQUENCE 241 AA; 25941 MW; 44EC9A0106AD1A68 CRC64;

Query Match 93.8%; Score 1086.5; DB 1; Length 241;
Best Local Similarity 91.6%; Pred. No. 1.9e-104;
Matches 206; Conservative 2; Mismatches 14; Indels 3; Gaps 2;

QY 1 IVGGYECTHSQAHQVSLNSGYHXCGLSLNXXVWVSAAHCKYKSLVRLGEGHHRVNEG 60
DB 20 IVGGYECTHSQAHQVSLNSGYHXCGLSLNXXVWVSAAHCKYKSLVRLGEGHHRVNEG 79
QY 61 TEQXISSSVXKHPNYSSYNINXNDIMLIKLTYPATLNOYHVAVALPTCAADATMCTVSG 120
DB 80 TEQYISSSVIRHPNYSSYNINNDIMLIKLTYPATLNOYHVAVALPTCAADATMCTVSG 139
QY 121 WGNMTSSVXDGLKQLXLPILSHADCANSGYCGMTQSMFCAGYLEGGKDCOGDSGGP 180
DB 140 WGNMTSSVADGDKLQCLSLPLSHADCANSGYCGMTQSMFCAGYLEGGKDCOGDSGGP 198
QY 181 VVCGVLTQGVVSWGYGCAERDPGVYAKVXVLSGWVRDTPMXY 225
DB 199 VVCGVLTQ--GVVSWGYGCAERDPGVYAKVXVLSGWVRDTPMXY 241

RESULT 2
TRYX GADMO STANDARD; PRT; 241 AA.
AC Q91041;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DB Trypsin X precursor (EC 3.4.21.4).
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.
OX NCBI_TaxID=8049;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pyeloric caeca;
RX MEDLINE=94039130; PubMed=8223632;
RA Gudmundsdottir A., Gudmundsdottir E., Oskarsson S., Bjarnason J.B.,
RA Eakin A.E., Craik C.S.;
RT "Isolation and characterization of cDNAs from Atlantic cod encoding
RT two different forms of trypsinogen."
RL Eur. J. Biochem. 217:1091-1097(1993).
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; X76887; CAAS4215.1; -.
DR HSBP; P00763; LDPO.
DR MEROPS; S01.151; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
DR Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
KW Multigene family.
FT SIGNAL 1 13 POTENTIAL.
FT PROPEP 14 19 ACTIVATION PEPTIDE.
FT CHAIN 20 241 TRYPSIN X.
FT ACT SITE 59 59 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 103 103 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 195 195 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 26 155 BY SIMILARITY.
FT DISULFID 44 60 BY SIMILARITY.
FT DISULFID 128 228 BY SIMILARITY.
FT DISULFID 135 201 BY SIMILARITY.
FT DISULFID 166 180 BY SIMILARITY.
FT DISULFID 191 215 BY SIMILARITY.
FT SITE 189 189 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 241 AA; 25976 MW; 853D7C26BCAF9DD7 CRC64;

Query Match 93.7%; Score 1085.5; DB 1; Length 241;
Best Local Similarity 91.6%; Pred. No. 2.4e-104;
Matches 206; Conservative 2; Mismatches 14; Indels 3; Gaps 2;

QY 1 IVGGYECTHSQAHQVSLNSGYHXCGLSLNXXVWVSAAHCKYKSLVRLGEGHHRVNEG 60
DB 20 IVGGYECTHSQAHQVSLNSGYHXCGLSLNXXVWVSAAHCKYKSLVRLGEGHHRVNEG 79
QY 61 TEQXISSSVXKHPNYSSYNINXNDIMLIKLTYPATLNOYHVAVALPTCAADATMCTVSG 120
DB 80 TEQYISSSVIRHPNYSSYNINNDIMLIKLTYPATLNOYHVAVALPTCAADATMCTVSG 139
QY 121 WGNMTSSVXDGLKQLXLPILSHADCANSGYCGMTQSMFCAGYLEGGKDCOGDSGGP 180
DB 140 WGNMTSSVDDGDKLQCLNLPILSHADCANSGYCGMTQSMFCAGYLEGGKDCOGDSGGP 198
QY 181 VVCGVLTQGVVSWGYGCAERDPGVYAKVXVLSGWVRDTPMXY 225
DB 199 VVCGVLTQ--GVVSWGYGCAERDPGVYAKVXVLSGWVRDTPMXY 241

RESULT 3
TRY1 SALSA STANDARD; PRT; 242 AA.
AC P35031;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DB Trypsin I precursor (EC 3.4.21.4).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=96035908; PubMed=7556223;
RA Mate R., Lorenz J.B., Smals A.O., Torrisen K.R.;
RT "Molecular cloning and characterization of anionic and cationic
RT variants of trypsin from Atlantic salmon."
RL Eur. J. Biochem. 232:677-685(1995).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.82 ANGSTROMS).

```

RA Smalas A.O., Hordvik A.;  
 RT "Structure determination and refinement of benzamidine-inhibited  
 RT trypsin from the North Atlantic salmon (Salmo salar) at 1.82-A  
 RL Acta Crystallogr. D 49:318-330(1993).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.82 ANGSTROMS).  
 RX MEDLINE=95148588; PubMed=7846025;  
 RA Smalas A.O., Heimstad E.S., Hordvik A., Willassen N.P., Male R.;  
 RT "Cold adaptation of enzymes: structural comparison between salmon and  
 RL bovine trypsins.";  
 RN Proteins 20:149-166(1994).  
 RP CC -|- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
 CC -|- SUBCELLULAR LOCATION: Extracellular.  
 CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X70075; CAA49680.1; -;  
 DR EMBL; X70071; CAA49676.1; -;  
 DR EMBL; X70072; CAA49677.1; -;  
 DR PIR; S31775; S31775.  
 DR PIR; S31776; S31776.  
 DR PIR; S31777; S31777.  
 DR PDB; 2TBS; 30-APR-94.  
 DR PDB; 1BIT; 01-NOV-94.  
 DR MEROPS; S01.151; -;  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; Tryp\_SPC; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;  
 KW Multigene family; 3D-structure.  
 FT SIGNAL 1 15 POTENTIAL.  
 FT PROPEP 16 20 ACTIVATION PEPTIDE.  
 FT CHAIN 21 242 TRYPSIN I.  
 FT ACT\_SITE 60 60 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 104 104 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 196 196 CHARGE RELAY SYSTEM.  
 FT DISULFID 27 156  
 FT DISULFID 45 61  
 FT DISULFID 129 229  
 FT DISULFID 136 202  
 FT DISULFID 167 181  
 FT DISULFID 192 216  
 FT SITE 190 190 REQUIRED FOR SPECIFICITY (BY SIMILARITY).  
 FT VARIANT 33 33 T -> A (IN TRYPSIN IA/IB).  
 SQ SEQUENCE 242 AA; 25958 MW; 43F5642498067E5A CRC64;  
 Query Match 77.8%; Score 901.5; DB 1; Length 242;  
 Best Local Similarity 75.1%; Pred. No. 1.9e-85;  
 Matches 169; Conservative 15; Mismatches 38; Indels 3; Gaps 2;  
 QY 1 IVGGYECTKXSOAHQVLSNGYHXCGLSLNXXWVWSAAHCYKSVLRVLGHEHHRVNEG 60  
 DB 21 IVGGYECRAKYSOTHQVLSNGYHFCGSLNXXWVWSAAHCYKSVLRVLGHEHHRVNEG 80  
 QY 61 TEQXISSSSVKRRHNYSSNYKNDIMLKLTYPATLNQYHVALPTPCAADATMTCTVSG 120  
 DB 81 SEQFISSSSRVIRHNYSSNYKNDIMLKLTYPATLNQYHVALPTPCAADATMTCTVSG 140  
 QY 121 WGMTMSVVDGKLOXLPLPILSHADCANSGPGMITQSMFCAGYLEGGKDCSCQDSGGP 180

Db 141 WGMTMSSTADSNKLCQNLIPILSYSDCNNSY-PGMITNAMPCAGYLEGGKDCSCQDSGGP 199  
 QY 181 VVCGNGVLOGVWVGWGYGCAERDXPGVAKVXVLSGWRDITMAXY 225  
 Db 200 VVCGNELQ--GVWSGYGCAEPGNFGVAKVICFNDLSTWASY 242  
 RESULT 4  
 TRY2\_SALSA  
 ID TRY2\_SALSA STANDARD; PRT; 231 AA.  
 AC P35032;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Trypsin II precursor (EC 3.4.21.4) (Fragment).  
 OS Salmo salar (Atlantic salmon).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.  
 OX NCBI\_TaxID=8030;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=96035908; PubMed=7556223;  
 RA Male R., Lorens J.B., Smals A.O., Torrisen K.R.;  
 RT "Molecular cloning and characterization of anionic and cationic  
 RT variants of trypsin from Atlantic salmon.";  
 RL Eur. J. Biochem. 232:677-685(1995).  
 CC -|- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
 CC -|- SUBCELLULAR LOCATION: Extracellular.  
 CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X70073; CAA49678.1; -;  
 DR PIR; S31778; S31778.  
 DR HSP; P35031; IBIT.  
 DR MEROPS; S01.258; -;  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR SMART; SM00020; Tryp\_SPC; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;  
 KW Multigene family.  
 FT NON\_TER 1 1  
 FT SIGNAL <1 4 POTENTIAL.  
 FT PROPEP 5 9 ACTIVATION PEPTIDE.  
 FT CHAIN 10 231 TRYPSIN II.  
 FT ACT\_SITE 49 49 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 93 93 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 185 185 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 16 145 BY SIMILARITY.  
 FT DISULFID 34 50 BY SIMILARITY.  
 FT DISULFID 118 218 BY SIMILARITY.  
 FT DISULFID 125 191 BY SIMILARITY.  
 FT DISULFID 156 170 BY SIMILARITY.  
 FT DISULFID 181 205 BY SIMILARITY.  
 FT SITE 179 179 REQUIRED FOR SPECIFICITY (BY SIMILARITY).  
 SQ SEQUENCE 231 AA; 24823 MW; C54A1CAPE74FAE18 CRC64;  
 Query Match 77.2%; Score 893.5; DB 1; Length 231;  
 Best Local Similarity 74.1%; Pred. No. 1.2e-84;  
 Matches 168; Conservative 16; Mismatches 38; Indels 3; Gaps 2;  
 QY 1 IVGGYECTKXSOAHQVLSNGYHXCGLSLNXXWVWSAAHCYKSVLRVLGHEHHRVNEG 60

Db 10 IVGGYCKAYSQPHVSLNSGYHFCGSLNENWVVAACHYQSRVVEVRLGHNQVTEG 69  
QY 61 TEQXISSSVXVRHPNYSYNYNDIMLIKLTYPATLNQYHVALPTECAADATMCTVSG 120  
Db 70 SEQFISSSVIRHPNYSYNYNDIMLIKLTYPATLNQYHVALPTECAADATMCTVSG 129  
QY 121 WGMTSSVVDGDKLOXLKLPILSHADCANSGYGMITQSMFCAGYLEGGKDCQGDSSGP 180  
Db 130 WGMTSSVVDGDKLOXLKLPILSHADCANSGYGMITQSMFCAGYLEGGKDCQGDSSGP 188  
QY 181 VVNCVGLQGVVSWGVCARDPGVYAKVXVLSGWVRDTMA 225  
Db 189 VVNCVGLQ--GVVSWGVCARDPGVYAKVXVLSGWVRDTMA 231

RESULT 5  
TRYP\_SQUAC STANDARD; PRT; 229 AA.  
AC P00764;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Trypsin precursor (EC 3.4.21.4).  
OS Squalus acanthias (Spiny dogfish).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
OC Elasmobranchii; Squala; Squaloidae; Squalidae; Squalus.  
OX NCBI\_TaxID=7797;  
RN [1]  
RP SEQUENCE OF 8-229.  
RX MEDLINE=75146445; PubMed=1092332;  
RA Titani K., Ericsson L.H., Neurath H., Walsh K.A.;  
RT "Amino acid sequence of dogfish trypsin."  
RL Biochemistry 14:1358-1366(1975).  
RN [2]  
RP SEQUENCE OF 1-21.  
RA Hermanson M.A., Tye R.W., Reeck G.R., Neurath H., Walsh K.A.;  
RT "Comparison of the amino terminal sequences of bovine, dogfish, and  
lungfish trypsins."  
RL FEBS Lett. 14:222-224(1971).  
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
CC -1- SUBCELLULAR LOCATION: Extracellular.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
DR PIR; A00950; TRDP.  
DR HSSP; P00763; IDPO.  
DR MEROPS; S01.151; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_Src; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen.  
FT PROPEP 1 7  
FT CHAIN 8 229  
FT ACT\_SITE 47 47  
FT ACT\_SITE 91 91  
FT ACT\_SITE 183 183  
FT DISULFID 14 143  
FT DISULFID 32 48  
FT DISULFID 116 216  
FT DISULFID 123 189  
FT DISULFID 154 168  
FT DISULFID 179 203  
FT SITE 177 177  
FT VARIANT 109 109 L -> P.  
SQ SEQUENCE 229 AA; 24591 MW; E83B83C5AD72FCE4 CRC64;

Query Match 65.5%; Score 758.5; DB 1; Length 229;  
Best Local Similarity 62.8%; Pred. No. 98-71; 57; Indels 3; Gaps 2;  
Matches 140; Conservative 23; Mismatches 57;

QY 1 IVGGYECTXHSQAHVSLNSGYHFCGSLNENWVVAACHYQSRVVEVRLGHNQVTEG 60  
Db 8 IVGGYECTXHSQAHVSLNSGYHFCGSLNENWVVAACHYQSRVVEVRLGHNQVTEG 67  
QY 61 TEQXISSSVXVRHPNYSYNYNDIMLIKLTYPATLNQYHVALPTECAADATMCTVSG 120  
Db 68 DETYIDSSVIRHPNYSYNYNDIMLIKLTYPATLNQYHVALPTECAADATMCTVSG 127  
QY 121 WGMTSSVVDGDKLOXLKLPILSHADCANSGYGMITQSMFCAGYLEGGKDCQGDSSGP 180  
Db 128 WGMTSSVVDGDKLOXLKLPILSHADCANSGYGMITQSMFCAGYLEGGKDCQGDSSGP 186  
QY 181 VVNCVGLQGVVSWGVCARDPGVYAKVXVLSGWVRDTMA 223  
Db 187 VVNCVGLQ--GVVSWGVCARDPGVYAKVXVLSGWVRDTMA 227

RESULT 6  
TRY3\_SALSA STANDARD; PRT; 238 AA.  
AC P35033;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Trypsin III precursor (EC 3.4.21.4) (Fragment).  
OS Salmo salar (Atlantic salmon).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.  
OX NCBI\_TaxID=8030;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Pancreas;  
RX MEDLINE=96035908; PubMed=7556223;  
RA Male R., Lorens J.B., Smals A.O., Torrisen K.R.;  
RT "Molecular cloning and characterization of anionic and cationic  
variants of trypsin from Atlantic salmon."  
RL Eur. J. Biochem. 232:677-685(1995).  
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
CC -1- SUBCELLULAR LOCATION: Extracellular.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

CC EMBL; X70074; CAA49679.1; -.  
CC PIR; S31779; S31779.  
CC HSSP; P00763; IDPO.  
CC MEROPS; S01.151; -.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR SMART; SM00020; Tryp\_Src; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;  
MW Multigene family.  
FT NON\_TER 1 1  
FT SIGNAL <1 7  
FT PROPEP 8 15  
FT CHAIN 16 238  
FT ACT\_SITE 55 55  
FT ACT\_SITE 99 99  
FT ACT\_SITE 192 192  
FT DISULFID 22 152  
FT DISULFID 40 56  
FT DISULFID 124 225  
FT POTENTIAL.  
FT ACTIVATION PEPTIDE.  
FT TRYPSIN III.  
FT CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT BY SIMILARITY.



```

FT DISULFID 131 198 BY SIMILARITY.
FT DISULFID 163 177 BY SIMILARITY.
FT DISULFID 188 212 BY SIMILARITY.
FT SITE 186 186 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 238 AA: 25389 MW; AE79980E893023 CRC64;

Query Match 65.3%; Score 756; DB 1; Length 238;
Best Local Similarity 63.8%; Pred. No. 1.7e-70;
Matches 143; Conservative 26; Mismatches 51; Indels 4; Gaps 3;

QY 1 IVGGYECTHSQAHQVSLNSGVCGLSLNXXVVSAAHCVKSLVRLVRLGHHIRVNEG 60
DQ 16 IVGGYECRKNSAYQSALQSGVHFCGSLSSVTVVVSAAHCVKSRIOVRLGHHIRVNEG 75
QY 61 TEQXISSSVYKRPNNYSVYNXNDIMLIKLTXPATLNQVVAVALPTECAADATCTVSG 120
DQ 76 TEQFIDSVKVMPPSYNSKNLNDIMLIKLSFASLNSVSTVALPSSCASSGTRCLVSG 135
QY 121 MONTMSSVXD-GDKLQXLPLSHADCANISYGPQMITSQMFAGYLEGGKDCQCGDSGG 179
DQ 136 MGNLSGSSNYPDTLRCLDLPILSSSSCNAY-PGQITSNMFACGFMWEGKDCQCGDSGG 194
QY 180 PWCNGVLQGVWGVGVCARDPGVYAKVXVLSGWRDTWA 223
DQ 195 PWCNGQLQ--GVWSGYGCAQRNKGVTYKVNYSWISWTMS 236

RESULT 7
TRVL RAT STANDARD; PRT; 246 AA.
AC P00762;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin I, anionic precursor (EC 3.4.21.4) (Pretrypsinogen I).
GN Trfi.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RC STRAIN=Sprague-Dawley;
RA MEDLINE=82265624; PubMed=6896710;
RA McDonald R.J., Stary S.J., Swift G.H.;
RT "Two similar but nonallelic rat pancreatic trypsinogens. Nucleotide
sequences of the cloned cDNAs.";
RL J. Biol. Chem. 257:9724-9732(1982).
RN [2]
RC SEQUENCE FROM N.A.
RA MEDLINE=85054880; PubMed=6094547;
RA Craik C.S., Choo Q.L., Swift G.H., Quinto C., McDonald R.J.,
RA Rutter W.J.;
RT "Structure of two related rat pancreatic trypsin genes.";
RL J. Biol. Chem. 259:14255-14264(1984).
RN [3]
RC X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RA MEDLINE=87292123; PubMed=3112942;
RA Sprang S., Standing T., Fletterick R.J., Stroud R.M., Finer-Moore J.,
RA Xuong N.-H., Hamlin R., Rutter W.J., Craik C.S.;
RT "The three-dimensional structure of Asn102 mutant of trypsin: role of
Asp102 in serine protease catalysis.";
RL Science 237:905-909(1987).
CC -|- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -|- SUBCELLULAR LOCATION: Extracellular.
CC -|- MISCELLANEOUS: THIS SEQUENCE REPRESENTS THE PRECURSOR OF THE MAJOR
CC FORM OF TRYPSIN PRODUCED BY THE ADULT PANCREAS.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: V01273; CAA24580.1; -
DR EMBL: J00778; AAA98518.1; -
DR PIR: A00948; TRRT1
DR PDB: 1TRM; 15-JUL-93.
DR PDB: 2TRM; 16-JUL-88.
DR PDB: 1BRA; 30-APR-94.
DR PDB: 1BRB; 31-JUL-94.
DR PDB: 1BRC; 31-MAY-94.
DR MEROPS: S01.094; -
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser protease_Try.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; Tryp_SPC; 1.
DR PROSITE: PS02040; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
KW Multigene family; 3D-structure.
FT SIGNAL 1 15
FT PROPEP 16 23
FT CHAIN 24 246
FT ACT_SITE 63 63
FT ACT_SITE 107 107
FT ACT_SITE 200 200
FT DISULFID 30 160
FT DISULFID 48 64
FT DISULFID 132 233
FT DISULFID 139 206
FT DISULFID 171 185
FT DISULFID 196 220
FT SITE 194 194
FT STRAND 25 25
FT STRAND 28 29
FT TURN 32 33
FT TURN 36 37
FT STRAND 38 42
FT STRAND 46 54
FT TURN 55 56
FT STRAND 57 60
FT HELIX 62 64
FT STRAND 70 73
FT STRAND 77 77
FT TURN 78 79
FT STRAND 86 95
FT TURN 97 98
FT STRAND 100 100
FT TURN 101 104
FT STRAND 105 105
FT TURN 106 106
FT STRAND 109 113
FT STRAND 127 127
FT TURN 135 136
FT STRAND 138 143
FT STRAND 157 157
FT STRAND 159 165
FT HELIX 168 174
FT TURN 176 178
FT TURN 181 182
FT STRAND 183 186
FT TURN 189 190
FT STRAND 194 194
FT TURN 197 198
FT TURN 200 201
FT STRAND 203 206
FT TURN 207 208
FT STRAND 209 216
FT TURN 223 224
FT STRAND 227 231

```

REQUIRED FOR SPECIFICITY.

```

FT HELIX 232 234
FT HELIX 236 244
FT TURN 245
SQ SEQUENCE 246 AA; 25959 MW; 6AFA0DAD11943FB5 CRC64;

Query Match 65.2%; Score 755; DB 1; Length 246;
Best Local Similarity 63.8%; Pred. No. 2.2e-70;
Matches 143; Conservative 25; Mismatches 52; Indels 4; Gaps 3;

QY 1 IVGGYECTXHSQAQVSLNSGYHXCGLINXVWVSAHCYKSVLRVLGEHHRVNEG 60
DB 24 IVGGYCTPEHSVPQVSLNSGYHFCGSLINDQWVWSAHCYKSRQVRLGEHHRVNEG 83
QY 61 TEQXISSSVXRPYSSNYXNDIMLIKLTXPATLNOYVHAVALPTECAADATMCTVSG 120
DB 84 DEQPINAAKIIXHPNYSSTLNDIMLIKLSPPVKLNARVAPVALPSACAPAGTQCLISG 143
QY 121 WGMTMSS-VXDGDGLQXKLPLILSHADCANSGYGMITQSMFCAGYLEGKDCSCQDSSG 179
DB 144 WGMTLSNGVNPDLLOQVDAPVLSQACEAAY-PGEITSSMICVGFLEGGKDCSCQDSSG 202
QY 180 PWCNGVLQGVVWSGVCGERDXPGVYAKVXVLSGWRDTMA 223
DB 203 PWCNGQLQ--GIVSWGVCALPDPNPGVYTKVCNPFVGMIOQTIA 244

RESULT 8
TRY2 XENLA STANDARD; PRT; 244 AA.
AC P70059;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin precursor (EC 3.4.21.4).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=87066713; PubMed=3641189;
RA Stevenson B.J., Hagenbuechle O., Wellauer P.K.;
RT "Sequence organisation and transcriptional regulation of the mouse
elastase II and trypsin genes".
RL Nucleic Acids Res. 14:8307-8330(1986).
CC -|- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -|- SUBCELLULAR LOCATION: Extracellular.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; U72330; AAB17274.1; -
DR HSSP; P00763; IDPO.
DR MEROPS; S01.258; -
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS0240; TRYPSPIN_DOM; 1.
DR PROSITE; PS00134; TRYPSPIN_HIS; 1.
DR PROSITE; PS00135; TRYPSPIN_SER; 1.
DR Hydrolase; Serine protease; Digestion; Zymogen; Signal;
KW Multigene family.
FT SIGNAL 1 15 BY SIMILARITY.
FT PROPEP 16 21 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 22 244 TRYPSPIN.
FT ACT_SITE 61 61 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 105 105 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 198 198 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 28 158 BY SIMILARITY.
FT DISULFID 46 62 BY SIMILARITY.
FT DISULFID 130 231 BY SIMILARITY.
FT DISULFID 137 204 BY SIMILARITY.
FT DISULFID 169 183 BY SIMILARITY.
FT DISULFID 194 218 BY SIMILARITY.
FT SITE 192 192 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 244 AA; 26079 MW; C63F29CB3300B323 CRC64;

Query Match 63.5%; Score 735; DB 1; Length 244;
Best Local Similarity 61.6%; Pred. No. 2.5e-68;
Matches 138; Conservative 33; Mismatches 49; Indels 4; Gaps 3;

QY 1 IVGGYECTXHSQAQVSLNSGYHXCGLINXVWVSAHCYKSVLRVLGEHHRVNEG 60
DB 22 IVGGFTCAKNAVYQVSLNAGYHFCGSLINSQWVSAHCYKSRQVRLGEHHRVNEG 81
QY 61 TEQXISSSVXRPYSSNYXNDIMLIKLTXPATLNOYVHAVALPTECAADATMCTVSG 120
DB 82 TEQFIDSQKVIKHPNYSNRNLNDIMLIKLTARLSANIOSVPLPSACAGTNCCLISG 141
QY 121 WGMTMSSVXD-GDKLQXKLPLILSHADCANSGYGMITQSMFCAGYLEGKDCSCQDSSG 179
DB 142 WGMTLSNGVNPDLLOQVDAPVLSQACEAAY-PGEITSSMICVGFLEGGKDCSCQDSSG 200
QY 180 PWCNGVLQGVVWSGVCGERDXPGVYAKVXVLSGWRDTMA 223
DB 201 PWCNGQLQ--GIVSWGVCALPDPNPGVYTKVCNPFVGMIOQTIS 242

RESULT 9
TRY2 MOUSE STANDARD; PRT; 246 AA.
AC P07146;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin II, anionic precursor (EC 3.4.21.4) (Pretrypsinogen II).
GN TRY2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/J;
RA Stevenson B.J., Hagenbuechle O., Wellauer P.K.;
RT "Sequence organisation and transcriptional regulation of the mouse
elastase II and trypsin genes".
RL Nucleic Acids Res. 14:8307-8330(1986).
CC -|- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -|- SUBCELLULAR LOCATION: Extracellular.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X04574; CAA28243.1; -
DR EMBL; X04577; CAA28245.1; -
DR PIR; B25528; B25528.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.064; -
DR MGD; MGI:102759; Try2.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.

```



```

FT CONFLICT 88      V -> I (IN REF. 1).
SQ SEQUENCE 246 AA; 26228 MW; ABD3630809AE6606 CRC64;

Query Match      62.7%; Score 726; DB 1; Length 246;
Best Local Similarity 61.6%; Pred. NO. 2.1e-67;
Matches 138; Conservative 25; Mismatches 57; Indels 4; Gaps

QY 1 IVGGYECTXHSQAHQVSLNSGVYHKCGGSLINXKXVWVSAAHCVKSVLVRVRLGBHHIRVNEG 60
DB 24 IVGGYTCQBSNVPVQVSLNSGVYHFCGSLINQWVWVSAAHCVKSRIOVRLGEHNINVLGG 83
QY 61 TEGXISSSVXRHPNYSYNIXNDIMLIKLTXPATLINQVYHVALPTECAADATMCTVSG 120
DB 84 NEQFVNAAKIIRHPNEDRXTLNNDIMLIKLSGPVKLNARVATVALPSSCAPAGTQCLISG 143
QY 121 WGNMTMS-VYDGDKLOXLPLILSHADCANSGYPGMITOSMCAGVLEGGKDSQCQSDSG 179
DB 144 WGNLTSSGVNEPDLQCLDAPLLPQADCEASY-PGKITDNMVCVGFLEGGKDSQCQSDSG 202
QY 180 PIVCVNGVLQGVVSGYGCABRDXPGVYAKXVVLSGWRDTMA 223
DB 203 PIVCNGELQ-GIVSWGYGALPDNFGVYTKVCNVYVDWIQDTIA 244

RESULT 12
TRY2 BOVIN
ID TRY2_BOVIN STANDARD; PRT; 247 AA.
AC Q29463;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin, anionic precursor (EC 3.4.21.4).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Holstein-Friesian; TISSUE=Pancreas;
RX MEDLINE=91065383; PubMed=1701147;
RA Le Huouon I., Wicker C., Guilloreau P., Touleec R., Puigserver A.;
RT "Isolation and nucleotide sequence of cDNA clone for bovine
RT pancreatic anionic trypsinogen. Structural identity within the
RT trypsin family.";
RL Eur. J. Biochem. 193:767-773(1990).
CC -! CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -! SUBCELLULAR LOCATION: Extracellular.
CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC between the Swiss Institute of Bioinformatics and the EMBL Outstac-
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commu-
CC entities requires a license agreement (See http://www.isb-sib.ch/ann-
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X54703; CAA38513.1; -.
CC HSSP; P00763; LDPO.
CC MEROPS; S01_258; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR SMART; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal.
KW SIGNAL
FT PROPEP 16 23 POTENTIAL.
FT CHAIN 24 247 ACTIVATION PEPTIDE.
FT CHAIN 24 247 TRYPSIN, ANIONIC.

```

FT ACT\_SITE 63 63 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 107 107 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 200 200 CHARGE RELAY SYSTEM.  
 FT DISULFID 30 160 BY SIMILARITY.  
 FT DISULFID 48 64 BY SIMILARITY.  
 FT DISULFID 132 233 BY SIMILARITY.  
 FT DISULFID 139 286 BY SIMILARITY.  
 FT DISULFID 171 185 BY SIMILARITY.  
 FT DISULFID 196 220 BY SIMILARITY.  
 FT SITE 194 194 REQUIRED FOR SPECIFICITY.  
 SQ SEQUENCE 247 AA; 26289 MW; 50A070495A7731DB CRC64;

Query Match 62.3%; Score 721; DB 1; Length 247;  
 Best Local Similarity 60.7%; Pred No. 7e-67;  
 Matches 136; Conservative 30; Mismatches 54; Indels 4; Gaps 3;

QY 1 IVGGVECTXHQHVSLSNGVHKGSLNXXVVSAAHCYKSVLRGLGEHHIRVNEG 60  
 Db 24 IVGGYCAENSVPYQVSLNAGYHFCGSLNDQWVVSAAHCYQVHIGRLGEYNDVLEG 83  
 QY 61 TEQIISSSVXRHPNYSSYNIXNDIMLIKTXPATLNQYVHAVALPTECAADATMCTVSG 120  
 Db 84 GEQFDASKIIRHPKYSSTWLDNILLIKLSTPAVINARVSTLLPSACASAGTECLISG 143  
 QY 121 WGMTSS-VYDGKLOXLPILSHADCANSPGMITQSMFCAGYLEGGKDCSQGDSGG 179  
 Db 144 WGMTSSGVNYPDLQCLVAPLLSHADCEASY-PGQITNNMICAGFLEGGKDCSQGDSGG 202  
 QY 180 PIVCNGVLQGVVSWMGVCAERDPPVAKVXVLSGWVDTMA 223  
 Db 203 PVACNGQLQ-GVSVMGYCAQKGGKPGVTKVCNVYDMIGETIA 244

RESULT 13  
 TRY3 RAT  
 ID TRY3 RAT STANDARD; PRT; 247 AA.  
 AC P08426;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Trypsin III, cationic precursor (EC 3.4.21.4) (Pretrypsinogen III).  
 GN TRY3.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87271609; PubMed=1607011;  
 RA Fletcher T.S., Alhadeff M., Craik C.S., Largman C.;  
 RT "Isolation and characterization of a cDNA encoding rat cationic  
 trypsinogen."  
 RL Biochemistry 26:3081-3086(1987).  
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg|-Xaa, Lys|-Xaa.  
 CC -!- SUBCELLULAR LOCATION: Extracellular.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M16624; AAA41985.1; -;  
 CC PIR; A27547; A27547.  
 CC DR HSPP; P00763; 1DPO.  
 CC DR MEROPS; S01.056; -;  
 CC DR InterPro; IPR001314; Chymotrypsin.  
 CC DR InterPro; IPR001254; Ser. protease\_Try.  
 CC Pfam; PF00089; trypsin; 1.  
 CC DR PRINTS; PRO0722; CHYMOTRYPSIN.

DR SMART; SM00020; Tryp\_SPC; 1.  
 DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_SER; 1.  
 DR PROSITE; PS00135; TRYPSIN\_HIS; 1.  
 KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;  
 KW Multigene family.  
 FT SIGNAL 1 15  
 FT PROPEP 16 24  
 FT CHAIN 25 247  
 FT ACT\_SITE 64 64  
 FT ACT\_SITE 108 108  
 FT ACT\_SITE 201 201  
 FT DISULFID 31 161  
 FT DISULFID 49 65  
 FT DISULFID 133 234  
 FT DISULFID 140 207  
 FT DISULFID 172 186  
 FT DISULFID 197 221  
 FT SITE 195 195  
 SQ SEQUENCE 247 AA; 26269 MW; D74892BAA584E4A8 CRC64;

Query Match 62.1%; Score 719; DB 1; Length 247;  
 Best Local Similarity 61.2%; Pred No. 1.1e-66;  
 Matches 137; Conservative 26; Mismatches 57; Indels 4; Gaps 3;

QY 1 IVGGVECTXHQHVSLSNGVHKGSLNXXVVSAAHCYKSVLRGLGEHHIRVNEG 60  
 Db 25 IVGGYCAENSVPYQVSLNAGYHFCGSLNDSQWVVSAAHCYKSVLRGLGEHHIRVNEG 84  
 QY 61 TEQIISSSVXRHPNYSSYNIXNDIMLIKTXPATLNQYVHAVALPTECAADATMCTVSG 120  
 Db 85 GEQFDASKIIRHPKYSSTWLDNILLIKLSTPAVINARVSTLLPSACASAGTECLISG 144  
 QY 121 WGMTSS-VYDGKLOXLPILSHADCANSPGMITQSMFCAGYLEGGKDCSQGDSGG 179  
 Db 145 WGMTSSGVNYPDLQCLVAPLLSHADCEASY-PGKITSNNMFCLEGGKDCSQGDSGG 203  
 QY 180 PIVCNGVLQGVVSWMGVCAERDPPVAKVXVLSGWVDTMA 223  
 Db 204 PIVCNGQLQ-GVSVMGYCAQKGGKPGVTKVCNVYDMIGETIA 245

RESULT 14  
 TRY1 XENLA  
 ID TRY1 XENLA STANDARD; PRT; 243 AA.  
 AC F19799;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Trypsin precursor (EC 3.4.21.4).  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=91007255; PubMed=2210372;  
 RA Shi Y.B., Brown D.D.;  
 RT "Developmental and thyroid hormone-dependent regulation of pancreatic  
 genes in Xenopus laevis."  
 RL Genes Dev. 4:1107-1113(1990).  
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg|-Xaa, Lys|-Xaa.  
 CC -!- SUBCELLULAR LOCATION: Extracellular.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M16624; AAA41985.1; -;  
 CC PIR; A27547; A27547.  
 CC DR HSPP; P00763; 1DPO.  
 CC DR MEROPS; S01.056; -;  
 CC DR InterPro; IPR001314; Chymotrypsin.  
 CC DR InterPro; IPR001254; Ser. protease\_Try.  
 CC Pfam; PF00089; trypsin; 1.  
 CC DR PRINTS; PRO0722; CHYMOTRYPSIN.

or send an email to license@isb-sib.ch).

CC EMBL; X53458; CAA37538.1; -

DR RA A35871; A35871.

DR HSP; P00763; LDPO.

DR MEROPS; S01151; -

DR InterPro; IPR001314; Chymotrypsin.

DR Pfam; PR00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR SMART; SM00020; TRYD\_SPC; 1.

DR PROSITE; PS00240; TRYPSIN\_DOM; 1.

DR PROSITE; PS00134; TRYPSIN\_HIS; 1.

DR PROSITE; PS00135; TRYPSIN\_SER; 1.

KW Hydrolase; Serine protease; Digestion; Zymogen; Signal;

KW Multigene family.

FT SIGNAL 1 15 BY SIMILARITY.

FT PROPEP 16 20 ACTIVATION PEPTIDE.

FT CHAIN 21 243 TRYPSIN.

FT ACT\_SITE 60 60 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 104 104 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 197 197 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT DISULFID 27 157 BY SIMILARITY.

FT DISULFID 45 61 BY SIMILARITY.

FT DISULFID 129 230 BY SIMILARITY.

FT DISULFID 136 203 BY SIMILARITY.

FT DISULFID 168 182 BY SIMILARITY.

FT DISULFID 193 217 BY SIMILARITY.

FT SITE 191 191 REQUIRED FOR SPECIFICITY (BY SIMILARITY).

SQ SEQUENCE 243 AA; 25492 MW; C5B8345A8B3F8031 CRC64;

Query Match 62.0%; Score 718; DB 1; Length 243;

Best Local Similarity 61.2%; Pred. No. 1.4e-66;

Matches 137; Conservative 31; Mismatches 52; Indels 4; Gaps 3;

Qy 1 IVGGVETXHSQAHQVSLNSGYHXCGLSLINXWVSAACHYKSVLRVRLGHEHIRVNEG 60

Db 21 IIGGATKAKSSVPYIVSLNSGYHFCGSLITNQWVSAACHYKASIQVRLGHEHIALSEG 80

Qy 61 TEQXISSSVKVRHPNYSYNTXNDIMLIKLTXPATLNQYVHVALPTECAADATMCTVSG 120

Db 81 TEQPISSSKVIRHSGNSYTLNDIMLIKLSASPASNAVNTVPLPGSCSAAGTSCSLISG 140

Qy 121 WGNMTSSVXD-GDKLOXLPLSHADCANSGVPGMITQSMFCAGYLEGKDSQCGDSGG 179

Db 141 WGNLTSGNSYVLDLQCLNAPLITNAQCNQNSAY-PGEITANVICVGYMEGGKDSQCGDSGG 199

Qy 180 PVVCGVQLQGVGVSWGVCACERDHPGVYAKVXVLSGVWRDTMA 223

Db 200 PVVCGVQLQ--GVVSWGVCACERNYPGVYTKVCYNAMIQNTIA 241

RESULT 15

TRY2\_HUMAN

ID\_TRY2\_HUMAN STANDARD; PRT; 247 AA.

AC P07478;

DT 01-APR-1988 (Rel. 07, Created)

DT 01-APR-1988 (Rel. 07, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Trypsin II precursor (EC 3.4.21.4) (Anionic trypsinogen).

GN PRS52 OR TRY2 OR TRYP2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=86221712; PubMed=3011602;

RA Emi M., Nakamura Y., Ogawa M., Yamamoto T., Nishide T., Mori T.,

RA Matsubara K.;

RT "Cloning, characterization and nucleotide sequences of two cDNAs

RL encoding human pancreatic trypsinogens.";

RL Gene 41:305-310(1986).

or send an email to license@isb-sib.ch).

CC EMBL; M27602; AAA61232.1; -

DR PIR; B25852; B25852.

DR HSP; P00763; LDPO.

DR MEROPS; S01258; -

DR Genew; HGNC:9483; PRSS2.

DR MIM; 601564; -

DR InterPro; IPR001314; Chymotrypsin.

DR InterPro; IPR001254; Ser protease\_Try.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR SMART; SM00020; TRYD\_SPC; 1.

DR PROSITE; PS00240; TRYPSIN\_DOM; 1.

DR PROSITE; PS00134; TRYPSIN\_HIS; 1.

DR PROSITE; PS00135; TRYPSIN\_SER; 1.

KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;

KW Multigene family.

FT SIGNAL 1 15 ACTIVATION PEPTIDE.

FT PROPEP 16 23 TRYPSIN II.

FT CHAIN 24 247 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 63 63 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 107 107 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 200 200 BY SIMILARITY.

FT DISULFID 30 160 BY SIMILARITY.

FT DISULFID 48 64 BY SIMILARITY.

FT DISULFID 171 185 BY SIMILARITY.

FT DISULFID 196 220 BY SIMILARITY.

FT SITE 194 194 REQUIRED FOR SPECIFICITY (BY SIMILARITY).

SQ SEQUENCE 247 AA; 26488 MW; 82B0F41EB9E3D5DB CRC64;

Query Match 61.3%; Score 710; DB 1; Length 247;

Best Local Similarity 60.3%; Pred. No. 9.5e-66;

Matches 135; Conservative 31; Mismatches 54; Indels 4; Gaps 3;

Qy 1 IVGGVETXHSQAHQVSLNSGYHXCGLSLINXWVSAACHYKSVLRVRLGHEHIRVNEG 60

Db 24 IVGGVICEENSVPQVSLNSGYHFCGSLISEQWWSAGHCYKSRIQVRLGHEHIEVLEG 83

Qy 61 TEQXISSSVKVRHPNYSYNTXNDIMLIKLTXPATLNQYVHVALPTECAADATMCTVSG 120

Db 84 NEQFINAAKIRHPKYNSTRDLNDILIKLSAPVAVNSRVSAISLPTAPPAGTESLISG 143

Qy 121 WGNMTSSVXD-GDKLOXLPLSHADCANSGVPGMITQSMFCAGYLEGKDSQCGDSGG 179

Db 144 WGNLTSGADYDPELQCLDAPVLSQACEASY-PGKITNNMFCVGLFEGGKDSQCGDSGG 202

Qy 180 PVVCGVQLQGVGVSWGVCACERDHPGVYAKVXVLSGVWRDTMA 223

Db 203 PVVSGNELQ--GIVSWGVCACAKRNPGVYTKVYVNDIMKDTIA 244

Search completed: February 12, 2003, 10:23:21

Job time : 31.2239 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2003, 10:03:55 ; Search time 151.119 Seconds  
(without alignments)  
198.395 Million cell updates/sec

Title: US-10-036-371-8

Perfect score: 1158

Sequence: 1 IVGGVECTXSHQHQVSLNS.....GVYAKVXVLSGWVDTNAY 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq 101002:\*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1130	97.6	225	22 AAB31579	Amino acid sequenc
2	886	76.5	240	23 ABB04644	Engraulis japonicu
3	883	76.3	241	23 ABB04645	Engraulis japonicu
4	755	65.2	246	19 AAM46260	Human amyloid beta
5	727	62.8	247	21 AAY78974	Canine cationic tr
6	715	61.7	247	21 AAB35701	Human trypsin hL a
7	710	61.3	247	23 AAU87693	Human pancreatic t
8	707	61.1	247	18 AAW08475	Porcine trypsinog
9	706	61.0	246	21 AAY78975	Canine anionic try
10	702	60.6	223	23 ABB83322	Partial trypsin se

11	702	60.6	224	20 AAY31160	Human trypsin seri
12	702	60.6	225	22 AAB98503	Human trypsin seri
13	702	60.6	230	20 AAW93488	Human TRY1 trypsin
14	702	60.6	247	21 AAB21321	Human trypsinogen
15	701	60.5	223	20 AAW81767	Bovine TRYP peptid
16	701	60.5	223	21 AAY69973	TRYP protein. Bos
17	701	60.5	224	15 AAR53637	Bovine trypsin. B
18	701	60.5	229	21 AAY77494	Bovine trypsinogen
19	701	60.5	231	21 AAY91926	Recombinant trypsi
20	701	60.5	231	22 AAB80953	Bovine met-phe-try
21	701	60.5	233	21 AAY91925	Trypsinogen analog
22	700	60.4	230	15 AAR53638	Bovine trypsinogen
23	692	59.8	241	21 AAB21316	Human trypsinogen
24	691	59.7	240	19 AAW57740	Trypsinogen-like p
25	665	57.4	247	9 AAP81243	Human spleen tryps
26	665	57.4	247	16 AAR82703	Human pancreatic t
27	500	43.2	238	20 AAY08894	Chimeric serine pr
28	494	42.7	256	23 AAU79390	Novel human kallik
29	494	42.7	320	23 AAE19166	Human protease, PR
30	494	42.7	320	23 AAU82732	Amino acid sequenc
31	466	40.2	151	22 AAU68927	Human protease dom
32	460	39.7	237	21 AAB21305	Human KLK-L6 prote
33	460	39.7	251	21 AAB21306	Human KLK-L6 prote
34	451	38.9	250	21 AAB21325	Human TLSP... Homo
35	451	38.9	250	21 AAY99390	Human PRO1279 (UNQ
36	451	38.9	250	22 ABB50479	Human secreted pro
37	451	38.9	250	22 AAU12424	Human PRO1279 poly
38	451	38.9	250	22 AAB66139	Protein of the inv
39	451	38.9	250	23 AAB661816	Prostate cancer-as
40	451	38.9	250	23 ABB95526	Human angiogenesis
41	451	38.9	250	23 ABB84920	Human PRO1279 prot
42	451	38.9	250	23 AAY93684	Human PRO protein,
43	451	38.9	281	20 AAY42440	CASB12 polypeptide
44	451	38.9	282	20 AAY42439	CASB12 amino acid
45	451	38.9	282	21 AAB11712	Human serine prote

#### ALIGNMENTS

RESULT 1

AAB31579

ID AAB31579 standard; peptide; 225 AA.

XX AAB31579;

XX 20-APR-2001 (first entry)

XX Amino acid sequence of cod trypsin isozymes.

XX Fish; serine proteinase; pain; acute inflammation; chronic inflammation;  
 KW arthritis; inflamed joint; bursitis; osteoarthritis; septic arthritis;  
 KW rheumatoid arthritis; juvenile rheumatoid arthritis; fibromyalgia;  
 KW systemic lupus erythematosus; phlebitis; tendinitis; rash; psoriasis;  
 KW acne; eczema; facial seborrheic eczema; foreskin infection;  
 KW athlete's foot; fissulae infection; ulcer; navel infection;  
 KW scar; kelloid; boil; wart; allergic itch; hemorroid; wound;  
 KW fungal infection; autoimmune disease.

XX Gadus callarias.

XX Key Location/Qualifiers

PH Misc-difference 9 /label= Lys, Arg

FT Misc-difference 24 /label= Tyr, Phe

FT Misc-difference 32 /label= Lys, Gln

FT Misc-difference 33 /label= Asp, Gln

FT Misc-difference 64 /label= Tyr, Phe

FT Misc-difference 71

```
FT      /label= Ile, Xaa
FT      /note= "Xaa is any amino acid or absent"
```

FT	Misc-difference 82	/label= Asp, Asn
FT	Misc-difference 92	/label= Lys, Glu
FT	Misc-difference 129	/label= Ala, Asp
FT	Misc-difference 136	/label= Val, Cys
FT	Misc-difference 138	/label= Asn, Ser
FT	Misc-difference 204	/label= His, Asn
FT	Misc-difference 212	/label= Met, Val
FT	Misc-difference 224	/label= Asn, Ser

PN WO200078332-A2.

28-DEC-2000.

15-JUN-2000: 2000WO-IS000005.

XX  
PR 18-JUN-1999: 99IS-0005086.

XX  
PA (BJAR/) BJARNASON J B.

XX PT Bjarnason JB:

XX  
WPT: 2001-091493/10

XX Fish serine proteinase, useful as a cosmetic, medicament for treating  
 PT eczema, psoriasis, arthritis, and in the manufacture of the medicament  
 PT for treating, preventing pathogenic diseases involving receptor  
 PT mediated binding -

PS Disclosure: Page 13; 38pp; English.

The specification describes a fish serine proteinase. The proteinases are useful as medicaments, for treating and preventing a disease in a human or an animal such as pain, acute inflammation, chronic inflammation, arthritis, inflamed joints, bursitis, osteoarthritis, rheumatoid arthritis, juvenile rheumatoid arthritis, septic arthritis, fibromyalgia, systemic lupus erythematosus, phlebitis, tendinitis, rash, psoriasis, acne, eczema, facial seborrhoeic eczema, eczema of the hands, face or neck, foreskin infections, athlete's foot, fistulae infections, infected topical ulcers, navel infections in newborns, wrinkles, scars, keloids, boils, warts and allergic itch, hemorrhoids, wounds, wound infections, wounds from burns, fungal infections and immunological and autoimmune diseases. They are also useful for removing dead or peeling skin from otherwise healthy skin, and for treating or preventing a disease in which pathogenesis is caused by bacteria, virus, fungus, parasite or a protozoan or a receptor mediated binding is involved. The present sequence represents the amino acid sequence of cod trypsin isozymes.

XX	Sequence	225 AA;
SO		

Query Match	97.6%;	Score 1130;	DB 22;	Length 225;
Best Local Similarity	100.0%;	Pred. No. 7.2e-96;		
Matches 225;	Conservative 0;	Mismatches 0;	Indels 0;	

0v 1 IVGGYECTXHSAHOVSLSNGYHXC GGSLINXXWVWSAAHCYKSVLRVRLGEHHIRVNEG 60

1 TVGGVECTYHISOAHOVS!NSGVYXCGGSLNXXWVWSAAHCYKSVLRVRLGEHHIRVNEG 60

61 TEQYSSSSSVRHPNVSSYNTXNDIMI.IKLTXPATLNOYVHAVALPTECAADATMCTVSG 120

61 TEQYSSSSVYRHPNYSSNTXNDIMI.KI.TXPA.TI.NOYVHAVALPTECAADATMCTVSG 120

121 WCNTMSSIVDCDKI.QYI.YI.PI.I.SHADCANSGPGMTTOSMFCAGYLEGGKDSCOGDSGGP 180

[illegible]

Db 121 WGN TMSSVXD GDKLQXLXLPILSHADCAN SYGPGMITQSMFCAGYLEGGK DSCQGDSSGP 180

QY 181 VVCNGVLQGVGVSWGYGCAERDXPGVYAKVXVLSGWVRDTMAXY 225

Db 181 VVCNGVLQGVGVSWGYGCAERDXPGVYAKVXVLSGHWVRTMAXY 225

## RESULT 2

ABB04644

XX

AC ABB04644;

DT 04-MAR-2002 (first entry)

DB Engraulis japonicus trypsinogen (aTry I) SEQ ID NO:1.

protease; trypsinogen; barding; Japanese anchovy; fish sauce.

OS Engraulis japonicus.

AA  
PN  
JP2001269173-A.

02-OCT-2001.

XX PF 24-MAR-2000: 2000JP-0084302.

XX  
PR 24-MAR-2000: 2000JP-0084302.XX  
PA  
(NTSB ) JAPAN TOBACCO INC.

XX WPT: 2002-078276/11

DR N-PSDB; ABA04345.  
vv

PT A new DNA sequence -

PS Claim 1; Fig 7; 32pp; Japanese.

The present sequence represents a trypsinogen, which is a protease (I) isolated from *Engraulis japonicus* (also called *Engraulis japonica* or Japanese anchovy). The present invention also describes: (1) a DNA encoding (I), or encoding a protease consisting of an amino acid sequence in which part of the amino acid residue is replaced, inserted or deleted in the amino acid sequence encoded by the above DNA and having a bio-activity substantially same as (I); (2) an expression vector in which the above DNA is recombined; (3) producing a sardine-derived protease in which a host cell transformed by the above expression vector is cultured and (I) is recovered; (4) a protease containing substantially no other protein derived from fish; and (5) the preparation of a fish sauce in a short period while inhibiting the generation of an unpleasant smell compared to a case where the protease prepared by the above method is not added in which at least one of a fish or a shellfish selected from the group consisting of Clupeidae order is immersed in an aqueous solution containing salts in as high salt concentration as about 8% to 24% and the protease prepared by the above method is added to it and fermented for about 1 to 11 months. The method is used for the preparation of a fish sauce in a short period.

AA	Sequence	240 AA;
SO		

Query Match 76.5%; Score 886; DB 23; Length 240;  
Best Local Similarity 74.9%; Pred. No. 1.8e-73;  
Matches 167; Conservative 14; Mismatches 38; Indels

1 TVGGVECTYHSHOAHOVSLNSGYHXCGGSLINXXWVWSAAHCYKSVLRVRLGEHHIRVNEG 60

20 TVGGVECOAHSOPHTVSI.NSGYHFCGGSLVNENWVWSAAHCYKSRVEVRLGEHHIGQEN 79

61 TEQYSSSSSVRHPNYSSNYXNDIMLIKLTXPATLNOYVHAVALPTECAADATMCTVSG 120

86 TEQRIKESBVIKPHQVSSYNIINQVMI.TKI.STPATI.NOYVOPVALPSRCASAGTMCLVAG 139





```

QY 61 TEQXISSSVXRHPNYSYNIKNDIMLIKLTXPATLNOYVHAVALPTECAADATMCTVSG 120
DB 84 DEQFINAAKIIRKPNYSSTWLNNDIMLIKLSPPVKLNARVAPVAFPSACAPAGTQCLISG 143
QY 121 WGNMTSS-VXDGGKQLXLPILSHADCANSGPGMITQSMFCAGYLEGGKDSQCGDSGG 179
DB 144 WNTLNGVNNPDLQCDAPVLSQADCEAA-Y-PGEITSSMICVGFLEGGKDSQCGDSGG 202
QY 180 PIVCNGVLQGVVSWGYGCAERDXPGVYAKVXVLSGWRDUMA 223
DB 203 PIVCNGQLQ--GIVSWGYGALPDNPGVYTKVCNPFVGIQDTIA 244

RESULT 5
AA78974
ID AAY78974 standard; protein; 247 AA.
XX
AC AAY78974;
XX
DT 05-JUN-2000 (first entry)
XX
DE Canine cationic trypsinogen amino acid sequence.
XX
KW Cationic trypsinogen; dog; monoclonal antibody production; detect;
KW diagnose; pancreatitis; pancreatic cancer; renal insufficiency;
KW extrapancreatic hyposecretion.
XX
OS Canis familiaris.
XX
PN WO200009739-A1.
XX
PD 24-FEB-2000.
XX
PF 09-AUG-1999; 99WO-JP04299.
XX
PR 10-AUG-1998; 98JP-0236609.
XX
PR 10-MAR-1999; 99JP-0063990.
XX
PA (FUJY ) FUJI YAKUHIN KOGYO KK.
XX
PI Waritani T, Ashida Y, Yamada T;
XX
WPI; 2000-206018/18.
XX
CC Anti-canine trypsin monoclonal antibody, useful as reagent for quick
CC and accurate detection and quantitation of trypsin and/or trypsin-like
CC immunoreactants in various forms in diagnosis e.g. of pancreatic
CC diseases -
CC
PS Claim 3; Page 63-64; 67pp; Japanese.
CC
CC This sequence represents the canine cationic trypsinogen amino acid
CC sequence. The protein was isolated from the canine pancreas. The
CC invention relates to monoclonal antibodies with specificity against
CC canine trypsin, or canine trypsin-related substances. The antibodies are
CC highly specific and can be used as reagent for quick and accurate
CC detection and quantitation of canine trypsin and canine trypsin-like
CC immunoreactants in various forms. The antibodies can be used in the
CC diagnosis of diseases such as pancreatitis, pancreatic cancer, renal
CC insufficiency and extrapancreatic hyposecretion.
CC
SQ Sequence 247 AA;
Query Match 62.8%; Score 727; DB 21; Length 247;
Best Local Similarity 61.6%; Pred. No. 7, 1e-59;
Matches 138; Conservative 31; Mismatches 51; Indels 4; Gaps 3;

QY 1 IVGGYECTXHSQAQVSLNSGYHXCGLSLINXXVWVSAACHYKSVLRVLGHHIRVNEG 60
DB 24 IVGGYTCENSVYPQVSLNAGYHFCGLSLISDQWVWVSAACHYKSVLRVLGHHIRVNEG 83
QY 61 TEQXISSSVXRHPNYSYNIKNDIMLIKLTXPATLNOYVHAVALPTECAADATMCTVSG 120

```

```

DB 84 NEQFINSAKIRHPNYSNWIILNDIMLIKLSPPAVLNARVATISLPACAAAPGTQCLISG 143
QY 121 WGNMTSSVXD-GDKLQXLKLPILSHADCANSGPGMITQSMFCAGYLEGGKDSQCGDSGG 179
DB 144 WNTLSSGTNYPELLQCLDAPILTOACEASY-PGQITENMICAGFLEGGKDSQCGDSGG 202
QY 180 PIVCNGVLQGVVSWGYGCAERDXPGVYAKVXVLSGWRDUMA 223
DB 203 PIVCNGELQ--GIVSWGYGCAQKNKPGVYTKVCNPFVDWIQSTIA 244

RESULT 6
AAB35701
ID AAB35701 standard; Protein; 247 AA.
XX
AC AAB35701;
XX
DT 14-FEB-2001 (first entry)
XX
DE Human trypsin hL amino acid sequence.
XX
KW Human; trypsin hL; serine protease; lung disease model animal.
XX
OS Homo sapiens.
XX
PN JP2000253887-A.
XX
PD 19-SEP-2000.
XX
PF 11-MAR-1999; 99JP-0065337.
XX
PR 11-MAR-1999; 99JP-0065337.
XX
PA (TTPH-) TT PHARMA KK.
XX
WPI; 2000-658814/64.
DR N-PSDB; AAC66182.
XX
PT Novel gene encoding a serine protease and its protein used to screen
PT for serine protease inhibitors and to prepare lung disease animal
PT models -
XX
PS Claim 2; Page 3-4; 17pp; Japanese.
XX
CC Nucleotide sequence AAC66182 encodes human trypsin hL AAB35701, a serine
CC protease. The invention relates to the human hL gene and protein
CC sequences, to a recombinant vector containing the nucleotide sequence,
CC and a host cell containing the vector. Human trypsin hL can be used for
CC screening for serine protease inhibitors, in the preparation of a lung
CC disease model animal, and for the development of an index marker of lung
CC diseases caused by an anti-trypsin hL antibody.
XX
SQ Sequence 247 AA;
Query Match 61.7%; Score 715; DB 21; Length 247;
Best Local Similarity 61.6%; Pred. No. 8, 9e-58;
Matches 138; Conservative 23; Mismatches 59; Indels 4; Gaps 3;

QY 1 IVGGYECTXHSQAQVSLNSGYHXCGLSLINXXVWVSAACHYKSVLRVLGHHIRVNEG 60
DB 25 IVGGYTCORNALPQVSLNSGYHFCGLSLISQWVWVSAACHYKSVLRVLGHHIRVNEG 84
QY 61 TEQXISSSVXRHPNYSYNIKNDIMLIKLTXPATLNOYVHAVALPTECAADATMCTVSG 120
DB 85 GEQPIDAAKIIRHPNYSNANTYNDIMLIKLTAAATLSRVSTVALPRSCPSAGTRCLVSG 144
QY 121 WGNMTSSVXD-GDKLQXLKLPILSHADCANSGPGMITQSMFCAGYLEGGKDSQCGDSGG 179
DB 145 WNTLSSGTNYPSLLQCLDAPVLSDSCTSSY-PGKITSNMFCLEGGKDSQCGDSGG 203
QY 180 PIVCNGVLQGVVSWGYGCAERDXPGVYAKVXVLSGWRDUMA 223
DB 204 PIVCNGQLQ--GVVSWGYGCAQKPGVYTKVCYVNVWIOQTIA 245

```

RESULT 7  
AAU87693  
ID AAU87693 standard; Protein; 247 AA.  
XX  
AC AAU87693;  
XX  
DT 21-MAY-2002 (first entry)  
XX  
DE Human pancreatic tumour protein #5.  
DE  
DE Human; pancreatic tumour protein; immune response; pancreatic cancer;  
KW development of cancer; cancer progression; cytostatic.  
OS  
OS Homo sapiens.  
XX  
PN WO200212331-A2.  
XX  
PD 14-FEB-2002.  
XX  
PF 06-AUG-2001; 2001WO-US24619.  
XX  
PR 07-AUG-2000; 2000US-223130P.  
PR 30-JAN-2001; 2001US-263447P.  
PR 15-MAY-2001; 2001US-291201P.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Pyle RA, Xu J, Kalos MD;  
XX  
DR WPI; 2002-241741/29.  
DR N-PSDB; ABK44208.  
XX  
DR Novel polynucleotide encoding pancreatic tumour polypeptides, useful in  
PT pharmaceutical compositions, e.g. vaccines, for treating pancreatic  
PT cancers -  
XX  
PS Claim 2; Page 164-165; 167pp; English.  
XX  
CC The present invention relates to the isolation of cDNA sequences  
CC (ABK44061-ABK44209) encoding human pancreatic tumour proteins. The  
CC polynucleotide sequences encoding human pancreatic tumour proteins are  
CC useful for stimulating an immune response in a patient and treating  
CC pancreatic cancer in a patient. A host cell that expresses these  
CC polynucleotides is useful for determining the presence of cancer in a  
CC patient. A composition comprising the polynucleotide, its encoded  
CC protein, or an antibody that binds to the protein may be used in the  
CC diagnosis, prevention and/or treatment of diseases, particularly  
CC pancreatic cancer. The sequences of the invention are also useful in  
CC pharmaceutical compositions, e.g. vaccines, for the diagnosis and  
CC treatment of pancreatic cancer. Such compositions may be useful for  
CC inhibiting the development of cancer in a patient, or as markers for  
CC the progression of cancer. The polynucleotide sequences may also be used  
CC as probes or primers for nucleic acid hybridisation assays.  
CC AAU87689-AAU87694 represent human pancreatic tumour proteins.  
XX  
SQ Sequence 247 AA;

Query Match 61.3%; Score 710; DB 23; Length 247;  
Best Local Similarity 60.3%; Pred. No. 2.6e-57;  
Matches 135; Conservative 31; Mismatches 54; Indels 4; Gaps 3;

QY 1 IVGGYECTXHSQAHQVSLNSGYHXCGLSLNXXVWSAAHCYKSVLRVLGHEHHIRVNEG 60  
DB 24 IVGGYICEENSVYQVSLNSGYHFCGSLSEQWVWSAGHCYKSRIVRLGHEHHIRVNEG 83  
QY 61 TEQXISSSVXRHPNYSYNIKNDIMLIKLTXPATLNOVYHVALEPTECAADATMCTVSG 120  
DB 84 NEQFINAAKIIIRHPKYNSTRNLNDILLKSSPAVINSRVSALISLTPAAPAAGTESLISG 143  
QY 121 WNTMSSVXD-GDKLQXLPLILSHADCANSGYGMITQSMFCAGYLEGGKSCQDSSG 179

DB 144 WNTLSSGADYPDELQCLDAPVLSQAECEASY-PGKITNNMFCVGLFEGGKSCQDSSG 202  
QY 180 PIVCNGVLQGVVWSWGYGCAERDYPGYAKVXVLSGWVDTMA 223  
DB 203 PIVSNGELQ--GIWSWGYGCAQNRPRGYTKVYNYVDWIKDTIA 244

RESULT 8  
AAW08475  
ID AAW08475 standard; Protein; 247 AA.  
XX  
AC AAW08475;  
XX  
DT 28-MAR-1997 (first entry)  
XX  
DE Porcine trypsinogen.  
XX  
KW Trypsinogen; trypsin; protease; filamentous fungus; Asperillus.  
XX  
OS Sus scrofa.  
XX  
PN WO9700316-A1.  
XX  
PD 03-JAN-1997.  
XX  
PF 10-JUN-1996; 96WO-DK00253.  
XX  
PR 16-JUN-1995; 95DK-0000693.  
XX  
PA (NOVO ) NOVO-NORDISK AS.  
XX  
PI Kjeidsen TB, Woldike HF;  
XX  
DR WPI; 1997-077523/07.  
DR N-PSDB; AAT49878.  
XX  
PT Prodn. of trypsin or trypsinogen in filamentous fungi - transformed  
PT with recombinant DNA comprising trypsinogen sequence fused to signal  
PT sequence, provides higher level of expression than other expression  
PT systems  
XX  
PS Example 2; Page 15; 28pp; English.  
XX  
CC An almost complete pig trypsinogen (AAW08475) lacking the very N-  
CC terminal end of the signal peptide is the product of a cDNA clone  
CC (AAT49878) isolated from a pig pancreas cDNA library. The cDNA clone  
CC can be fused to a signal sequence (e.g. from the native enzyme or  
CC from a specified amylase gene) and used for prodn. of recombinant  
CC trypsin in transformed host cells, esp. filamentous fungus, e.g.  
CC Aspergillus oryzae or Aspergillus niger. Levels of trypsin  
CC secreted by Aspergillus spp. are several-fold higher than those  
CC obtd. using other microbial systems.  
XX  
SQ Sequence 247 AA;

Query Match 61.1%; Score 707; DB 18; Length 247;  
Best Local Similarity 60.3%; Pred. No. 4.8e-57;  
Matches 135; Conservative 27; Mismatches 58; Indels 4; Gaps 3;

QY 1 IVGGYECTXHSQAHQVSLNSGYHXCGLSLNXXVWSAAHCYKSVLRVLGHEHHIRVNEG 60  
DB 25 IVGGYTCANSPYQVSLNSGSHFCGSLNSQWVWSAAHCYKSRIVRLGHEHHIRVNEG 84  
QY 61 TEQXISSSVXRHPNYSYNIKNDIMLIKLTXPATLNOVYHVALEPTECAADATMCTVSG 120  
DB 85 NEQFINAAKIIIRHPKYNSTRNLNDILLKSSPAVINSRVSALISLTPAAPAAGTESLISG 144  
QY 121 WNTMSSVXD-GDKLQXLPLILSHADCANSGYGMITQSMFCAGYLEGGKSCQDSSG 179  
DB 145 WNTKSGSSYPSLQCLKAPVLSDSSCASSY-PQITGNMTCVGLFEGGKSCQDSSG 203  
QY 180 PIVCNGVLQGVVWSWGYGCAERDYPGYAKVXVLSGWVDTMA 223

Wed Feb 12 11:59:46 2003

us-10-036-371-8.rag

Db 204 PIVCNGOLO--GIVSWGCGCAQKKEGVYTKVCNYYNWIQOTIA 245

RESULT 9

AA78975

ID AAY78975 standard; protein; 246 AA.

XX AC

XX AAY78975;

XX DT

XX 05-JUN-2000 (first entry)

XX DE Canine anionic trypsinogen amino acid sequence.

XX KW Anionic trypsinogen; dog; monoclonal antibody production; detect;

XX KW diagnose; pancreatitis; pancreatic cancer; renal insufficiency;

XX KW extrapancreatic hyposecretion.

XX OS

XX Canis familiaris.

XX PN WO200009739-A1.

XX PD

XX 24-FEB-2000.

XX PF

XX 09-AUG-1999; 99WO-JP04299.

XX PR

XX 10-AUG-1998; 98JP-0236609.

XX PR

XX 10-MAR-1999; 99JP-0063990.

XX PA (FUJY) FUJI YAKUHIN KOGYO KK.

XX PI Waritani T, Aehida Y, Yamada T;

XX DR WPI; 2000-206018/18.

XX PT Anti-canine trypsin monoclonal antibody, useful as reagent for quick

XX PT and accurate detection and quantitation of trypsin and/or trypsin-like

XX PT immunoreactants in various forms in diagnosis e.g. of pancreatic

XX PT diseases

XX PS Claim 3; Page 64-65; 67pp; Japanese.

XX CC This sequence represents the canine anionic trypsinogen amino acid

XX CC sequence. The protein was isolated from the canine pancreas. The

XX CC invention relates to monoclonal antibodies with specificity against

XX CC canine trypsin, or canine trypsin-related substances. The antibodies are

XX CC highly specific and can be used as reagent for quick and accurate

XX CC detection and quantitation of canine trypsin and canine trypsin-like

XX CC immunoreactants in various forms. The antibodies can be used in the

XX CC diagnosis of diseases such as pancreatitis, pancreatic cancer, renal

XX CC insufficiency and extrapancreatic hyposecretion.

XX SQ Sequence 246 AA;

Query Match 61.0%; Score 706; DB 21; Length 246;

Best Local Similarity 60.7%; Pred. No. 5.9e-57;

Matches 136; Conservative 28; Mismatches 56; Indels 4; Gaps 3;

Qy 1 IVGGYECTHSQAHQVSLNSGYHXCGLSLINXXVWVSAACHYKSVLRVLRGHHIRVNEG 60

Db 24 IVGGYCTSRNSVPYQVSLNSGYHFCGGLSLNSQWVWVSAACHYKSVLRVLRGHHIRVNEG 83

Qy 61 TEQXISSSSVXRHPNYSYNIKXNDIMLIKLTXPATLNOYVHAVALPTECAADATMCTVSG 120

Db 84 GQOFINAKIIRHPNRYNANTINDIMLIKLSPTATLNSRVSAIALPKSPAAAGTQCLISG 143

Qy 121 WGNMTSSVXD-GDKLQXLPLILSHADCANSGYPMITQSMFCAGYLEGGKDCSCQDSGG 179

Db 144 WGNQTSIGQNYPDVLCAPILSDSVCRNAY-PGQISSNMCLGYMEGGKDCSCQDSGG 202

Qy 180 PIVCNGVLOGVGVWSNGYGCARDXPGVYAKVXVLSGWRDTMA 223

Db 203 PIVCNGELQ--GVVSWAGCAQKKGKGVSPKVCYKYSWIQOTIA 244

RESULT 10

ABB83322

ID ABB83322 standard; protein; 223 AA.

XX AC

XX ABB83322;

XX DT

XX 30-AUG-2002 (first entry)

XX DE Partial trypsin sequence.

XX KW Trypsin; three dimensional structure; protein structure analysis;

XX KW serine protease.

XX OS

XX Unidentified.

XX PN US2002035434-A1.

XX PD

XX 21-MAR-2002.

XX PF

XX 23-JUL-2001; 2001US-0910054.

XX PR

XX 06-FEB-1992; 92JP-0021012.

XX PR

XX 11-DEC-1992; 92JP-0331703.

XX PR

XX 08-FEB-1993; 93US-0014867.

XX PA (FUJI) FUJITSU LTD.

XX PI Tomikawa M, Aikawa S, Matsuzawa F;

XX DR WPI; 2002-507172/54.

XX PT Analysis of three-dimensional structures by generating combination of

XX PT correspondence satisfying restriction condition, and calculating root

XX PT mean square distance between elements in the combination of

XX PT correspondence

XX PS Disclosure; Fig 38; 65pp; English.

XX CC The present invention relates to a method for analysis of three

XX CC dimensional structures. The method involves generating a combination of

XX CC correspondence satisfying a restriction condition between the elements

XX CC belonging to a first and second point sets from among all candidates for

XX CC the combination of correspondence, and calculating a root mean square

XX CC distance between the elements corresponding in the combination of

XX CC correspondence. The method is useful for analysing three-dimensional

XX CC structures of molecules, particularly proteins. To illustrate the

XX CC invention, the serine proteases trypsin and elastase were used. The

XX CC present sequence is a partial sequence for trypsin. The histidine and

XX CC serine active sites of trypsin and elastase were analysed using the

XX CC method of the invention.

XX SQ Sequence 223 AA;

Query Match 60.6%; Score 702; DB 23; Length 223;

Best Local Similarity 60.3%; Pred. No. 1.2e-56;

Matches 135; Conservative 29; Mismatches 56; Indels 4; Gaps 3;

Qy 1 IVGGYECTHSQAHQVSLNSGYHXCGLSLINXXVWVSAACHYKSVLRVLRGHHIRVNEG 60

Db 1 IVGGYCTCCANTPYQVSLNSGYHFCGGLSLNSQWVWVSAACHYKSVLRVLRGHHIRVNEG 60

Qy 61 TEQXISSSSVXRHPNYSYNIKXNDIMLIKLTXPATLNOYVHAVALPTECAADATMCTVSG 120

Db 61 NEQFISAKSVTHFESYNSNTLNNDIMLIKLSAASLNSRVASISLPTSCASAGTQCLISG 120

Qy 121 WGNMTSS--VXDGDKLQXLPLILSHADCANSGYPMITQSMFCAGYLEGGKDCSCQDSGG 179

Db 121 WGNYSKSGTSPDVLKCLAPILSDSSCKSAY-PGQITSNMFCAGYLEGGKDCSCQDSGG 179

Qy 180 PIVCNGVLOGVGVWSNGYGCARDXPGVYAKVXVLSGWRDTMA 223

Db 180 PIVCSGKQLQ--GIVSWGSGCAQKKNKPGVYTKVCNYSWIKOTIA 221



Wed Feb 12 11:59:46 2003

Db 181 PIVCNGQLQ--GVVSWGDCGCAQKNKFGVYTKVYVYVVKWIKNTIA 222

RESULT 13  
AAW93488  
ID AAW93488 standard; Protein; 230 AA.  
XX  
AC AAW93488;  
XX  
DT 11-JUN-1999 (first entry)  
XX  
DE Human TRYI trypsinogen variant protein.  
XX  
KW TRYI; trypsinogen; autocatalysis; cleavage; zymogenic precursor;  
KW protease; peptide hormone; therapeutic protein; treatment; coagulation.  
XX  
KW Homo sapiens.  
OS  
XX  
PN WO9910503-A1.  
XX  
PD 04-MAR-1999.  
XX  
PF 12-AUG-1998; 98WO-EP05094.  
XX  
PR 15-OCT-1997; 97EP-0117816.  
XX  
PR 22-AUG-1997; 97EP-0114513.  
XX  
PA (HOFF ) ROCHE DIAGNOSTICS GMBH.  
XX  
PI Bode W, Hopfner K, Huber R, Kopetzki E;  
XX  
XX WPI; 1999-204669/17.  
XX  
DR N-PSDB; AAX23298.  
XX  
XX Autocatalytically cleavable zymogenic protease precursors - useful  
XX for cleaving fusion proteins and for therapeutic uses  
XX  
XX Disclosure; Fig 1; 45pp; German.  
XX  
XX This invention describes a method where autocatalytically cleavable,  
XX zymogenic precursors of a protease (in this invention trypsinogen),  
XX have their natural cleavage site replaced by a non-natural,  
XX autocatalytically cleavable site. Such proteases are reagents for  
XX producing peptide hormones and other therapeutic proteins by cleavage  
XX of their fusion proteins, and some also have therapeutic activity  
XX themselves, e.g. thrombin for treatment of coagulation disorders. The  
XX proteases are produced simply and in high yield without the need to  
XX add another protease for cleavage, reducing costs and avoiding  
XX contamination of the final enzyme.  
XX  
XX SQ Sequence 230 AA;  
Query Match 60.6%; Score 702; DB 20; Length 230;  
Best Local Similarity 60.3%; Pred. No. 1.3e-56;  
Matches 135; Conservative 26; Mismatches 59; Indels 4; Gaps 3;  
QY 1 IVGGYECTHSAQHOVSLNSGYHXCGLSLNXXVWVSAACHYKSVLRVLRGHHIRVNEG 60  
DB 7 IVGGYCNCEENSPVQVSLNSGYHFCGGLSLNEQWVVSAGHCYKSRIQVRLGHNIEVLEG 66  
QY 61 TEQXISSSVVXRHPNYSYNIIXNDIMLIKLTXPATLNQYVHVALPTECAADATMCTVSG 120  
DB 67 NEQFINAAKTIIRHPQYDRKTLNNDIMLIKLSRAVINARVSTISLPTAPPATGTKCLISG 126  
QY 121 WGMTMSSVXD-GDKLQXKLXLPILSHADCANISYGPMTQSMFCAGYLEGGKDCSCGDSGG 179  
DB 127 WGMTASSGADYPDELQCLDAPVLVSQAKCEASY-PGKITSNMFCVGFLEGGKDCSCGDSGG 185  
QY 180 PIVCNGVLQGVVSWGYGCAERDXPGVYAKVXVLSGWVRDTMA 223  
DB 186 PIVCNGQLQ--GVVSWGDCGCAQKNKFGVYTKVYVYVVKWIKNTIA 227

RESULT 14  
AAB21321  
ID AAB21321 standard; Protein; 247 AA.  
XX  
AC AAB21321;  
XX  
DT 02-FEB-2001 (first entry)  
XX  
DE Human trypsinogen.  
XX  
KW Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6; trypsinogen;  
KW kallikrein-like protein; serine protease;  
KW cytostatic; cancer; prostate cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO200053776-A2.  
XX  
PD 14-SEP-2000.  
XX  
PF 09-MAR-2000; 2000WO-CA00258.  
XX  
PR 11-MAR-1999; 99US-0124260.  
XX  
PR 01-APR-1999; 99US-0127386.  
XX  
PR 21-JUL-1999; 99US-0144919.  
XX  
PA (MOUN ) MOUNT SINAI HOSPITAL.  
XX  
PI Yousef GM, Diamandis EP;  
XX  
XX WPI; 2000-587440/55.  
XX  
XX New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L  
XX protein mediated disorders, especially cancer. -  
XX  
XX Example 4; Fig 17; 184pp; English.  
XX  
XX The present sequence is human trypsinogen, a member of the serine  
XX protease family. Kallikreins and kallikrein-like proteins are a  
XX subgroup of the serine protease enzyme family. They catalyze the  
XX selective cleavage of specific polypeptide precursors to release  
XX peptides with potent biological activity. Nucleic acids encoding  
XX kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4, KLK-L5 and  
XX KLK-L6 have been isolated. The proteins are useful in the treatment,  
XX monitoring and diagnosis of cancers, especially prostate cancer. They  
XX can also be used to identify a substance that can associate with or  
XX mediate the biological activity of the proteins. Antibodies can  
XX be used to treat conditions mediated by the kallikrein-like proteins.  
XX  
XX SQ Sequence 247 AA;  
Query Match 60.6%; Score 702; DB 21; Length 247;  
Best Local Similarity 60.3%; Pred. No. 1.4e-56;  
Matches 135; Conservative 26; Mismatches 59; Indels 4; Gaps 3;  
QY 1 IVGGYECTHSAQHOVSLNSGYHXCGLSLNXXVWVSAACHYKSVLRVLRGHHIRVNEG 60  
DB 24 IVGGYCNCEENSPVQVSLNSGYHFCGGLSLNEQWVVSAGHCYKSRIQVRLGHNIEVLEG 83  
QY 61 TEQXISSSVVXRHPNYSYNIIXNDIMLIKLTXPATLNQYVHVALPTECAADATMCTVSG 120  
DB 84 NEQFINAAKTIIRHPQYDRKTLNNDIMLIKLSRAVINARVSTISLPTAPPATGTKCLISG 143  
QY 121 WGMTMSSVXD-GDKLQXKLXLPILSHADCANISYGPMTQSMFCAGYLEGGKDCSCGDSGG 179  
DB 144 WGMTASSGADYPDELQCLDAPVLVSQAKCEASY-PGKITSNMFCVGFLEGGKDCSCGDSGG 202  
QY 180 PIVCNGVLQGVVSWGYGCAERDXPGVYAKVXVLSGWVRDTMA 223  
DB 203 PIVCNGQLQ--GVVSWGDCGCAQKNKFGVYTKVYVYVVKWIKNTIA 244

RESULT 15

Search completed: February 12, 2003, 10:22:29  
Job time : 152.119 secs

**THIS PAGE BLANK (USPTO)**